October 1985 Vol. 64 No. 8

AT&T TECHNICAL JOURNAL

A JOURNAL OF THE AT&T COMPANIES

Auditory Modeling

Speech Recognition

Traffic

Laser Applications

Digital Radio

Input-Output Maps

AT&T TECHNICAL JOURNAL (ISSN 8756-2324) is published ten times each year by AT&T, 550 Madison Avenue, New York, NY 10022; C. L. Brown, Chairman of the Board; L. L. Christensen, Secretary. The Computing Science and Systems section and the special issues are included as they become available. Subscriptions: United States—1 year $35; foreign—1 year $45.

Payment for foreign subscriptions or single copies must be made in United States funds, or by check drawn on a United States bank and made payable to the AT&T Technical Journal, and sent to AT&T Bell Laboratories, Circulation Dept., Room 1E335, 101 J. F. Kennedy Pky, Short Hills, NJ 07078.

Back issues of the special, single-subject supplements may be obtained by writing to the AT&T Customer Information Center, P.O. Box 19901, Indianapolis, Indiana 46219, or by calling (800) 432-6600. Back issues of the general, multisubject issues may be obtained from University Microfilms, 300 N. Zeeb Road, Ann Arbor, Michigan 48106.

Printed in U.S.A. Second-class postage paid at Short Hills, NJ 07078 and additional mailing offices. Postmaster: Send address changes to the AT&T Technical Journal, Room 1E335, 101 J. F. Kennedy Pky, Short Hills, NJ 07078.

AT&T
TECHNICAL
JOURNAL

VOL. 64 OCTOBER 1985 NO. 8

AT&T Technical Journal
Vol. 64, No. 8, October 1985
Printed in U.S.A.

Almost-Periodic Response Determination for Models of the Basilar Membrane

By I. W. SANDBERG and J. B. ALLEN*

(Manuscript received April 8, 1985)

Electrical networks consisting of linear passive elements and many nonlinear resistors are often used to model the basilar membrane. The inputs to these networks are typically a sum of sinusoids switched on at $t = 0$, and the resulting quantities of interest because of their interpretation as analogs of experimental observables are the steady-state response components of a certain current and of certain voltages. In this paper, recently obtained mathematical results concerning the input-output representation of nonlinear systems are used to give, for the first time, a locally convergent expansion for all of the steady-state quantities of interest. Also given is a good deal of information concerning general properties of the expansion, and this establishes important properties of the nonlinear network's response. Of particular practical interest is a term in the expansion that contains a component whose frequency is $(2f_1 - f_2)$ when the network's input consists of a sum of two sinusoids, with frequencies f_1 and f_2. One of our main results is an explicit expression for this $(2f_1 - f_2)$ component.

I. INTRODUCTION

Electrical networks of the type shown in Fig. 1, together with sophisticated frequency-domain measurement techniques, play a central role in the modeling and analysis of the peripheral auditory system.[1-9] In the figure—which shows a one-dimensional lumped-element transmission-line model of the basilar membrane—the inductors and capacitors are linear, the box at the upper left contains

* Authors are employees of AT&T Bell Laboratories.

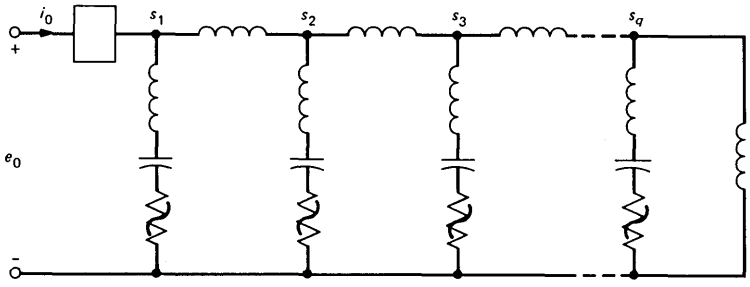


Fig. 1—Network model.

lumped elements, and, as indicated, the resistors are nonlinear. The voltage e_0 applied to the network is typically a finite sum of sinusoids (often a sum of just two sinusoids) switched on at some finite time that we take to be $t = 0$. The resulting quantities of interest, because of their interpretation as analogs of experimental observables, are the steady-state response components of the current i_0 and of one or more of the voltages $s_1, \cdots, s_q$.

In models of interest today the number q of nonlinear resistors is typically taken to be between 200 and 500. The resistors are assumed to be current controlled, with each current-voltage relationship often represented by the sum of linear and cubic terms.[1-3]

The purpose of this paper is to use recently obtained results[10] concerning the input-output representation of nonlinear systems to give, for the first time, an expansion for all of the steady-state quantities of interest in Fig. 1. The expansion is in terms of e_0 and is locally convergent. By this we mean that whenever the sum of the Fourier coefficients of e_0 is sufficiently small, and some reasonable additional conditions are met, the steady-state quantities exist and are given by the sum of the terms in the expansion, with each term dependent on the frequencies and Fourier coefficients of e_0. We emphasize that the expansion provides an exact representation of the response; it is not merely an approximation or a formal expansion whose convergence has not been proved. However, in this paper we do not give lower bounds on the *size* of the region of convergence. Questions of this type are the subject of ongoing studies.[11]

In Section II it will become clear that the terms in the expansion are defined by a certain recursive process. Of particular practical interest at the present time is the term we call the third-order term, which contains a component whose (radian) frequency is $(2\omega_1 - \omega_2)$ when e_0 consists of a sum of two sinusoids, one of frequency ω_1, and another of frequency ω_2. One of our main results is an explicit expression for this $(2\omega_1 - \omega_2)$ component, under some very reasonable assumptions.

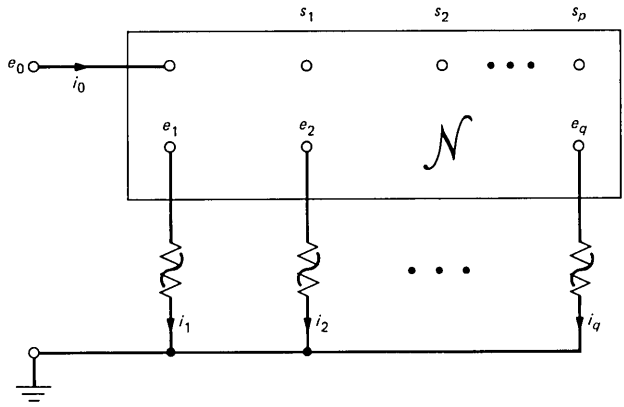


Fig. 2—More general network model.

II. EXISTENCE, PROPERTIES, AND EVALUATION OF THE STEADY-STATE QUANTITIES

2.1 Formulation

To enable attention to be more sharply focused on the concepts of importance to us, it is helpful to generalize our problem. Thus, we consider instead of Fig. 1 the network of Fig. 2, in which $\mathcal{N}$ is a linear time-invariant network and $s_1, \cdots, s_p$ are voltages in $\mathcal{N}$, measured with respect to the ground terminal, where p is any positive integer.

Let i and e, respectively, denote the transpose of the current and voltage row vectors $(i_1, \cdots, i_q)$ and $(e_1, \cdots, e_q)$. Assume that $\mathcal{N}$ has the representation

$$i(t) = \int_0^t h_a(t-\tau)e_0(\tau)d\tau + \int_0^t h_c(t-\tau)e(\tau)d\tau + u_1(t), \qquad t \geqslant 0, \quad (1)$$

in which h_a and h_c are $q \times 1$ and $q \times q$ matrix-valued impulse response functions and u_1 (which takes into account initial conditions) is a bounded continuous function that approaches zero as $t \to \infty$.* Similarly, let r stand for the transpose of the response $(s_1, \cdots, s_p, i_0)$ and suppose that there are $(p+1) \times 1$ and $(p+1) \times q$ matrix-valued impulse response functions h_d and h_b, respectively, for which

* We could have assumed that u_1 and the transient functions u_2 and u_3 to be introduced are all zero functions. However, we wish to establish that the steady-state responses are *robust* with respect to these functions in the strong sense that, under the conditions to be described, they are independent of them.

$$r(t) = \int_0^t h_d(t-\tau)e_0(\tau)d\tau + \int_0^t h_b(t-\tau)e(\tau)d\tau + u_2(t),$$

$$t \geq 0, \quad (2)$$

where u_2 is also a bounded continuous function that approaches zero as $t \to \infty$.

Each element of h_a, h_b, h_c, and h_d is assumed to be an absolutely integrable real-valued function on $[0, \infty)$ with possibly an impulse at $t = 0$. We use H_a to denote the Fourier transform of h_a, i.e.,

$$H_a(\omega) = \int_0^\infty h_a(t)e^{-j\omega t}dt, \qquad -\infty < \omega < \infty.$$

Similarly, H_b, H_c, and H_d stand for the Fourier transforms of h_b, h_c, and h_d, respectively. Of course $H_a(\omega)$, $H_b(\omega)$, $H_c(\omega)$, and $H_d(\omega)$ are also matrices. Notice that, from (1) and (2), each of these matrices has a natural transfer-function interpretation. For example, from (1) we see that the elements of H_a are the voltage-to-current transfer functions from the system input e_0 to the "inputs" i of the nonlinear resistors, when these resistors are replaced with short circuits.

The nonlinear resistors in Fig. 2 are assumed to be represented by

$$e_k(t) = R_k[i_k(t)], \qquad (k = 1, \cdots, q) \quad (3)$$

with each R_k an analytic function in some neighborhood of the origin of the complex plane, such that $R_k(z)$ is real when z is real, $R_k(0) = 0$, and $dR_k(z)/dz = 0$ at $z = 0$. (In particular, the R_k can be polynomials with real coefficients.) In Fig. 1 the nonlinear resistors typically have a relatively large linear part. These linear parts can be taken into account in Fig. 2 in $\mathcal{N}$. Using known properties of networks with positive elements, it is not difficult to show that the assumptions made above concerning h_a, h_b, h_c, h_d, u_1, and u_2 are satisfied for the network of Fig. 1 when put in the form of Fig. 2, as long as the linear part of each resistor has positive resistance, all linear elements are passive, the impedance of the two-terminal box is not zero at zero frequency, and each s_k in Fig. 2 is an s_k in Fig. 1.

2.2 Steady-state responses: properties and evaluation

We now assume that e_0 is given by

$$e_0(t) = \sum_{k=-\infty}^{\infty} a_k e^{j\omega_k t} + u_3(t), \qquad t \geq 0,$$

in which the sum of the $|a_k|$ is finite; $j = (-1)^{1/2}$; the ω_k are real; and u_3, like u_1 and u_2, is bounded, continuous, and approaches zero as $t \to \infty$. We do not require that the ω_k are multiples of some fixed constant.

Thus, the input is assumed to be the sum, for $t \geq 0$, of a so-called "almost-periodic" signal

$$\sum_{k=-\infty}^{\infty} a_k e^{j\omega_k t}, \qquad -\infty < t < \infty \tag{4}$$

and a transient part u_3. Although all almost-periodic signals have a generalized Fourier series of the form (4), the sum of the magnitudes of the Fourier coefficients need not be finite. We shall use AP to denote the subset of almost-periodic functions for which this sum *is* finite.

At this point we are able to state our main result, which is: Under the assumptions already discussed, and for $\sum_{k=-\infty}^{\infty} |a_k|$ as well as u_1, u_2, and u_3 sufficiently small,*

1. There are unique bounded functions i, e, and r that satisfy eqs. (1), (2), and (3), and (regarding uniqueness) a certain very reasonable neighborhood condition† concerning i,

2. There is a $(p + 1)$-vector-valued function r_{ss} defined on $(-\infty, \infty)$, with each of its $(p + 1)$ components belonging to (AP), such that

$$r(t) - r_{ss}(t) \to 0 \quad \text{as} \quad t \to \infty$$

(i.e., the response r approaches the steady state r_{ss} as $t \to \infty$), and

3. r_{ss} is independent of u_1, u_2, and u_3. It is given by

$$r_{ss}(t) = \sum_{m=1}^{\infty} [r_{ss}(t)]_m, \qquad -\infty < t < \infty, \tag{5}$$

in which the $[r_{ss}(\cdot)]_m$ are $(p + 1)$-vector-valued functions, with components belonging to AP, defined by

$$[r_{ss}(t)]_1 = \sum_{k=-\infty}^{\infty} H_d(\omega_k) a_k e^{j\omega_k t}$$

and

$$[r_{ss}(t)]_m = \sum_{k_1=-\infty}^{\infty} \cdots \sum_{k_m=-\infty}^{\infty} \mathscr{R}_m(\omega_{k_1}, \cdots, \omega_{k_m}) a_{k_1} \cdots a_{k_m} e^{j(\omega_{k_1}+\cdots+\omega_{k_m})t} \tag{6}$$

for $m \geq 2$, where the $\mathscr{R}_m(\omega_{k_1}, \cdots, \omega_{k_m})$, which depend on H_a, H_b, H_c, and the derivatives of the R_k at the origin, but not on the coefficients a_k, are defined by the recursive relations (10), (11), and (12) in the Appendix. The infinite sum in (5) converges uniformly in t.

Notice that a fundamental property of the class of network models

* By "small" for u_1, u_2, and u_3 is meant small in the reasonable sense of the $\mathscr{B}_0$ norm of Ref. 10, p. 692.

† The condition is simply that the function i must lie in a certain neighborhood of the origin. See the first of the two footnotes in the Appendix.

considered is that, with excitations as indicated, each component of any steady-state response r_{ss} belongs to AP. In particular, the r_{ss} are well behaved; any r_{ss} is continuous in t and has a Fourier series, and the Fourier series converges to $r_{ss}(t)$ for each t.

It is shown in the Appendix that the result described above follows from the main theorem in Ref. 10. In addition, bounds in Ref. 10, Section 2.4.3 show that the following can be added to 1 through 3.

4. There are positive constants α and β such that, with $([r_{ss}(t)]_m)_k$ the kth component of $[r_{ss}(t)]_m$,

$$\sum_{m=(M+1)}^{\infty} \max_k |([r_{ss}(t)]_m)_k| \leq \alpha \left(\beta \sum_{k=-\infty}^{\infty} |a_k|\right)^{(M+1)}, \qquad -\infty < t < \infty,$$

for any positive integer M [which provides useful information concerning the error in discarding all terms in (5) beyond the Mth].

2.3 The $(2\omega_1 - \omega_2)$ component of $[r_{ss}(t)]_3$

Each $[r_{ss}(t)]_m$ in (5) is of order m in the sense that the effect of multiplying all of the Fourier coefficients of e_0 by a constant λ is to cause $[r_{ss}(t)]_m$ to be replaced by $\lambda^m[r_{ss}(t)]_m$. Of particular interest in applications is an explicit expression for $T(\omega_1, \omega_2, a_1, a_2)$, the component at the frequency $(2\omega_1 - \omega_2)$ of the *third* order term in (5), when

$$e_0 = a_1 e^{j\omega_1 t} + a_{-1} e^{-j\omega_1 t} + a_2 e^{j\omega_2 t} + a_{-2} e^{-j\omega_2 t},$$

where a_{-1} and a_{-2} are the complex conjugates of a_1 and a_2, respectively, $0 < \omega_1 < \omega_2 < 2\omega_1$,* and $\alpha_k(l) = 0$ $(k = 1, \cdots, q)$ for $l = 2$, and where here and in the Appendix $\alpha_k(l)$ denotes $d^l R_k(z)/dz^l|_{z=0}$.

Under the condition on the α_k (2) indicated, the expression (12) for the $\mathscr{R}_m$ yields

$$\mathscr{R}_3(\omega_{k_1}, \omega_{k_2}, \omega_{k_3}) = \frac{1}{6} H_b(\omega_{k_1} + \omega_{k_2} + \omega_{k_3}) \text{diag}[\alpha_1(3), \cdots, \alpha_q(3)] \cdot \hat{\chi}[H_a(\omega_{k_1}), H_a(\omega_{k_2}), H_a(\omega_{k_3})],$$

where "diag" indicates a diagonal matrix and $\hat{\chi}[H_a(\omega_{k_1}), H_a(\omega_{k_2}), H_a(\omega_{k_3})]$ denotes the q-vector whose kth element is the product $[H_a(\omega_{k_1})]_k[H_a(\omega_{k_2})]_k[H_a(\omega_{k_3})]_k$ of kth elements for each k. Thus, using (6) with $m = 3$, $a_0 = 0$, and $a_k = 0$ for $|k| > 2$, as well as the observation that $(\omega_{k_1} + \omega_{k_2} + \omega_{k_3}) = (2\omega_1 - \omega_2)$ only if one of the ω_{k_l} is $-\omega_2$ and

* For ω_1 and ω_2 that meet these conditions, $(2\omega_1 - \omega_2)$ is not equal to ω_1, ω_2, $3\omega_1$, $3\omega_2$ or $(2\omega_2 - \omega_1)$, which are the only other positive frequencies at which $[r_{ss}(t)]_3$ can have components. However, higher-order terms of odd index *can* possess components at $(2\omega_1 - \omega_2)$. For example, $(\omega_{k_1} + \cdots + \omega_{k_5}) = (2\omega_1 - \omega_2)$ if $\omega_{k_1} = \omega_{k_2} = \omega_1$, $\omega_{k_3} = -\omega_2$, and $\omega_{k_4} + \omega_{k_5} = 0$.

the other two are equal to ω_1, it easily follows that the sum of the coefficients of $\exp[j(2\omega_1 - \omega_2)t]$ in $[R_{ss}(t)]_3$ is

$$\frac{1}{2} H_b(2\omega_1 - \omega_2)\text{diag}[\alpha_1(3), \cdots, \alpha_q(3)]\hat{\chi}[H_a(\omega_1), H_a(\omega_1), H_a(-\omega_2)]a_1^2 a_{-2}.$$

This shows that

$$T(\omega_1, \omega_2, a_1, a_2) = \text{Re}\{H_b(2\omega_1 - \omega_2)\text{diag}[\alpha_1(3), \cdots, \alpha_q(3)] \cdot \hat{\chi}[H_a(\omega_1), H_a(\omega_1), H_a(-\omega_2)]a_1^2 a_{-2}\exp[j(2\omega_1 - \omega_2)t]\},$$

where Re{ } stands for the real part of { }. Since H_a and H_b have a direct interpretation in terms of the structure of the network of Fig. 2, so does $T(\omega_1, \omega_2, a_1, a_2)$.

As a matter of convenience we have chosen to let $\mathcal{N}$ take into account the linear parts of the nonlinear resistors. We could have assumed instead that $\mathcal{N}$, without these linear parts, has sufficient damping that our conditions on h_a, h_b, h_c, h_d, u_1, and u_2 are satisfied. Under some very reasonable assumptions (see Ref. 10, comments on p. 694 concerning H.3), our expression for $T(\omega_1, \omega_2, a_1, a_2)$ would then explicitly exhibit its dependence on these linear parts, and this may be of interest in some cases. It can be shown, using a result in Ref. 10, Section 2.4.3, that the alternative expression for $T(\omega_1, \omega_2, a_1, a_2)$ that we would have obtained is

$$\text{Re}\{H_b(2\omega_1 - \omega_2)F(2\omega_1 - \omega_2)\text{diag}[\alpha_1(3), \cdots, \alpha_q(3)] \cdot \hat{\chi}[E(\omega_1)H_a(\omega_1), E(\omega_1)H_a(\omega_1), E(-\omega_2)H_a(-\omega_2)] \cdot a_1^2 a_{-2}\exp[j(2\omega_1 - \omega_2)t]\},$$

in which, with 1_q the identity matrix of order q,

$$F(\omega) = \{1_q - \text{diag}[\alpha_1(1), \cdots, \alpha_q(1)]H_c(\omega)\}^{-1},$$

and

$$E(\omega) = \{1_q - H_c(\omega)\text{diag}[\alpha_1(1), \cdots, \alpha_q(1)]\}^{-1}$$

(with both inverses existing for $-\infty < \omega < \infty$).

2.4 Discussion

In this paper we have derived and discussed a general expansion for the response of a cochlear model having a nonlinear membrane. The nonlinearities of the model take into account the membrane's nonlinear damping. Of particular interest is the third-order term in the expansion for the case described in Section 2.3, in which the input is a sum of sinusoids at frequencies ω_1 and ω_2. This term is the first term in the expansion that gives rise to a component at the frequency $(2\omega_1 - \omega_2)$.

The expression for the third-order term is seen to depend on two transfer-function matrices H_a and H_b, where H_b relates the output response vector r to the voltages across the resistors in Fig. 2 under the condition that e_0 is zero, and H_a relates the currents through the resistors to the input voltage e_0 under the condition that the resistors are replaced by short circuits.

In the expression for T, the transfer function H_b is evaluated only at $(2\omega_1 - \omega_2)$. The function H_b has the interpretation that it corresponds to a filter that alters the distortion products after their generation on the basilar membrane.

The terms $\alpha_1(3), \cdots, \alpha_q(3)$ are measures of the generator strength of the nonlinear distortion as a function of position, in the sense that each $\alpha_k(3)$ is proportional to the coefficient of the cubic term in the power series expansion of the resistor function R_k. Cubic nonlinearities have been used previously in basilar membrane models to model the generation of distortion products.

The transfer function H_a enters the expression for T in a particularly interesting way. Notice that any element of T, say the lth, is the real part of

$$\sum_{k=1}^{q} [H_b(2\omega_1 - \omega_2)]_{lk}\alpha_k(3)[H_a(\omega_1)]_k^2[H_a(-\omega_2)]_k a_1^2 a_{-2} e^{j(2\omega_1-\omega_2)t},$$

which is a linear combination of q terms with, so to speak, H_a appearing three times in each term, twice for ω_1 and once for ω_2.

REFERENCES

1. J. L. Hall, "Observations on a Nonlinear Model for Motion of the Basilar Membrane," in *Hearing Research and Theory*, Vol. 1, New York: Academic Press, 1981, pp. 1–61.
2. J. L. Hall, "Two-Tone Distortion Products in a Nonlinear Model of the Basilar Membrane," J. Acoust. Soc. Am., *56* (December 1974), pp. 1818–28.
3. J. W. Matthews, "Modeling Reverse Middle Ear Transmission of Acoustic Distortion Signals," in *Mechanics of Hearing*, ed. by E. de Boer and M. A. Viergever, Netherlands: Delft University Press, 1983, pp. 11–18.
4. E. Zwicker, "Cubic Difference Tone Level and Phase Dependence on Frequency and Level of Primaries," in *Psychophysical, Physiological and Behavioral Studies in Hearing*, ed. by van den Brink and Bilsen, Netherlands: Delft University Press, 1981.
5. J. L. Goldstein, "Auditory Nonlinearity," J. Acoust. Soc. Am., *41* (March 1967), pp. 676–89.
6. J. L. Goldstein, G. Buchsbaum, and M. Furst, "Compatibility Between Psychophysical and Physiological Measurements of Aural Combination Tones," J. Acoust. Soc. Am., *63* (February 1978), pp. 474–85.
7. J. L. Goldstein and N. Y. S. Kiang, "Neural Correlates of the Aural Combination Tones," Proc. IEEE, *56* (June 1968), pp. 981–92.
8. D. O. Kim, C. C. Molnar, and J. W. Matthews, "Cochlear Mechanics: Nonlinear Behavior in Two-Tone Response as Reflected in Cochlear-Nerve-Fiber Responses and in Ear-Canal Sound Pressure," J. Acoust. Soc. Am., *67* (May 1980), pp. 1704–21.
9. J. B. Allen and P. F. Fahey, "Nonlinear Behavior at Threshold Determined in the Auditory Canal and on the Auditory Nerve," in *Hearing-Physiological Bases and*

Psychophysics, ed. by R. Klinke and R. Hartman, New York: Springer-Verlag, 1983, pp. 128–34.
10. I. W. Sandberg, "Existence and Evaluation of Almost Periodic Steady-State Responses of Mildly Nonlinear Systems," IEEE Trans. Circuits Syst., *31*, No. 8 (August 1984), pp. 689–701.
11. I. W. Sandberg, "Criteria for the Global Existence of Functional Expansions for Input-Output Maps," AT&T Tech. J., *64*, No. 7 (September 1985), pp. 1639–58.

APPENDIX

Proof of the Main Steady-State Response Result; Recursive Relations for the $\mathscr{R}_m$

Theorem 3 of Ref. 10 would be directly applicable to the network governed by (1), (2), and (3) if $h_a(t-\tau)$, $h_b(t-\tau)$, $h_c(t-\tau)$, and $h_d(t-\tau)$ were *square* matrices of the *same* size. Since this condition is not met, we proceed to construct a suitable related set of system equations.

Let $n = (q + p + 2)$, and define K_a, K_b, K_c, and K_d to be the convolution operators associated with h_a, h_b, h_c, and h_d, respectively. Let v, x, y, and w be given by $v = (e_0, u_1, u_2)^{\text{tr}}$, $x = (i, x^{[p+2]})^{\text{tr}}$, $y = (e, y^{[p+2]})^{\text{tr}}$, and $w = (r, w^{[q+1]})^{\text{tr}}$, where "tr" denotes transpose, and $x^{[p+2]}$, $y^{[p+2]}$, and $w^{[q+1]}$ are unspecified vector-valued functions (on $t \geqslant 0$) of the indicated dimensions. Notice that v, x, y, and w are all n-vector valued. Finally, let $\eta_k (k = 1, \cdots, n)$ be the functions defined by $\eta_k = R_k (1 \leqslant k \leqslant q)$, with η_k equal to the zero function for $(q + 1) \leqslant k \leqslant n$.

Consider the equations

$$x = Av + Cy \tag{7}$$

$$w = Dv + By \tag{8}$$

$$y = Nx \tag{9}$$

in which by (9) we mean $y_k(t) = \eta_k[x_k(t)]$ for each k and t, and in which A, C, D, and B are given in partitioned form by

$$A = \begin{pmatrix} K_a & I(q) & Z(q, p+1) \\ Z(p+2, 1) & Z(p+2, q) & Z(p+2, p+1) \end{pmatrix},$$

$$C = \begin{pmatrix} K_c & Z(q, p+2) \\ Z(p+2, q) & Z(p+2, p+2) \end{pmatrix},$$

$$D = \begin{pmatrix} K_d & Z(p+1, q) & I(p+1) \\ Z(q+1, 1) & Z(q+1, q) & Z(q+1, p+1) \end{pmatrix},$$

and

$$B = \begin{pmatrix} K_b & Z(p+1, p+2) \\ Z(q+1, q) & Z(q+1, p+2) \end{pmatrix},$$

where, for any positive integers q and s, $I(q)$ denotes the identity operator on the space of q-vector valued functions on $t \geqslant 0$, and $Z(q, s)$ is the zero operator from the space of s-vector-valued functions on $t \geqslant 0$ into the corresponding space of functions whose values are of dimension q.

We see that if (7), (8), and (9) are satisfied, then (1), (2), and (3) are met, and that if the latter set of equations are satisfied and $x^{[p+2]}$, $y^{[p+2]}$, and $w^{[q+1]}$ are zero functions, then (7), (8), and (9) are satisfied. Using the fact that A, B, C, D, and N meet the conditions of Theorem 3 of Ref. 10, it follows from that theorem that statements 1 and 2 of Section 2.2 hold.* It also follows from the theorem that r_{ss} is independent of u_1, u_2, and u_3, and using the relation $w = (r, w^{[q+1]})^{\text{tr}}$, that $r_{ss}(t)$ can be written in the form (5) with the components of each $[r_{ss}(\cdot)]_m$ elements of AP, with

$$[r_{ss}(t)]_1 = \sum_{k=-\infty}^{\infty} H_d(\omega_k) a_k e^{j\omega_k t}$$

and each $[r_{ss}(t)]_m$ for $m \geqslant 2$ specified as follows (after some straightforward analysis involving partitioned matrices).

With $c_1, c_2, \cdots$ arbitrary n-vectors, and $\beta_1, \beta_2, \cdots$ arbitrary real numbers, let q-vector-valued functions $Q_1, Q_2, \cdots$ and $(p+1)$-vector-valued functions $P_2, P_3, \cdots$ be defined by $Q_1(c_1, \beta_1) = H_a(\beta_1)(c_1)_1$,

$$Q_m(c_1, \cdots, c_m, \beta_1, \cdots, \beta_m) = H_c(\beta_1 + \cdots + \beta_m) S_m$$

for $m \geqslant 2$, and

$$P_m(c_1, \cdots, c_m, \beta_1, \cdots, \beta_m) = H_b(\beta_1 + \cdots + \beta_m) S_m$$

for $m \geqslant 2$, in which†

$$S_m = \sum_{l=2}^{m} (l!)^{-1} \sum_{\substack{k_1+\cdots+k_l=m \\ k_j>0}} \text{diag}[\alpha_1(l), \cdots, \alpha_q(l)]$$

$$\cdot \hat{\chi}[Q_{k_1}(c_1, \cdots, c_{k_1}, \beta_1, \cdots, \beta_{k_1}), \cdots,$$

$$Q_{k_l}(c_{(m-k_l+1)}, \cdots, c_m, \beta_{(m-k_l+1)}, \cdots, \beta_m)],$$

$(c_1)_1$ is the first component of c_1,

$$\alpha_k(l) = \left. \frac{d^l R_k(z)}{dz^l} \right|_{z=0} \quad (k = 1, \cdots, q)$$

for each l, "diag" indicates a diagonal matrix, and $\hat{\chi}$ is defined by the

---

* The "neighborhood condition" of statement 1 is inherited from Ref. 10, part (iib) of Theorem 3 via the relationship between (1), (2), and (3) and (7), (8), and (9).

† In the expression for S_m, $\sum_{\substack{k_1+\cdots+k_l=m \\ k_j>0}}$ denotes a sum over all positive integers $k_1, \cdots, k_l$ that add to m.

condition that $\hat{\chi}[c_1, \cdots, c_l]$ is the q-vector with kth element $(c_1)_k \cdots (c_l)_k (1 \leq k \leq q)$. In terms of these P_m, we have

$$[r_{ss}(t)]_m = \sum_{k_1=-\infty}^{\infty} \cdots \sum_{k_m=-\infty}^{\infty} P_m(d_{k_1}, \cdots, d_{k_m}, \omega_{k_1}, \cdots, \omega_{k_m}) e^{j(\omega_{k_1}+\cdots+\omega_{k_m})t},$$

where $d_k = (a_k, 0, \cdots, 0)^{\text{tr}}$ for each k.

Observe that for any m and k with $1 \leq k \leq m$, each Q_m and each P_m is linear in c_k and independent of $(c_k)_l$ for $l \geq 2$. Thus, each $Q_m(c_1, \cdots, c_m, \beta_1, \cdots, \beta_m)$ is equal to $Q_m(u, \cdots, u, \beta_1, \cdots, \beta_m)(c_1)_1 \cdots (c_m)_1$, where $u = (1, 0, \cdots, 0)^{\text{tr}}$, and similarly for the P_m. Therefore, if $\mathcal{S}_m$ and $\mathcal{R}_m$ are defined by

$$\mathcal{S}_1(\beta_1) = H_a(\beta_1), \tag{10}$$

$$\mathcal{S}_m(\beta_1, \cdots, \beta_m) = H_c(\beta_1 + \cdots + \beta_m) \sum_{l=2}^{m} (l!)^{-1} \sum_{\substack{k_1+\cdots+k_l=m \\ k_j>0}} \text{diag}[\alpha_1(l), \cdots, \alpha_q(l)] \cdot \hat{\chi}[\mathcal{S}_{k_1}(\beta_1, \cdots, \beta_{k_1}), \cdots, \mathcal{S}_{k_l}(\beta_{(m-k_l+1)}, \cdots, \beta_m)] \tag{11}$$

for $m \geq 2$, and

$$\mathcal{R}_m(\beta_1, \cdots, \beta_m) = H_b(\beta_1 + \cdots + \beta_m) \cdot \sum_{l=2}^{m} (l!)^{-1} \sum_{\substack{k_1+\cdots+k_l=m \\ k_j>0}} \text{diag}[\alpha_1(l), \cdots, \alpha_q(l)] \cdot \hat{\chi}[\mathcal{S}_{k_1}(\beta_1, \cdots, \beta_{k_1}), \cdots, \mathcal{S}_{k_l}(\beta_{(m-k_l+1)}, \cdots, \beta_m)] \tag{12}$$

for $m \geq 2$, we have

$$[r_{ss}(t)]_m = \sum_{k_1=-\infty}^{\infty} \cdots \sum_{k_m=-\infty}^{\infty} \mathcal{R}_m(\omega_{k_1}, \cdots, \omega_{k_m}) a_{k_1} \cdots a_{k_m} e^{j(\omega_{k_1}+\cdots+\omega_{k_m})t}, \quad -\infty < t < \infty$$

for $m = 2, 3, \cdots$. This completes the Appendix.

AUTHORS

Jont B. Allen, B.S. (Electrical Engineering), 1966, University of Illinois, Urbana—Champaign; M.S. and Ph.D., University of Pennsylvania, Philadelphia, in 1968 and 1970, respectively; AT&T Bell Laboratories, 1970—. Mr. Allen is currently working in the areas of small room acoustics, dereverberation of speech signals, cochlear modeling, and digital signal processing. His main efforts have been directed toward modeling the cochlea.

Irwin W. Sandberg, B.E.E., 1955, M.E.E., 1956, and D.E.E., 1958, Polytechnic Institute of Brooklyn; AT&T Bell Laboratories, 1958—. Mr. Sandberg has been concerned with analysis of radar systems for military defense, synthesis and analysis of active and time-varying networks, with several fundamental studies of properties of nonlinear systems, and with some problems in communication theory and numerical analysis. His more recent interests have included compartmental models, the theory of digital filtering, global implicit-function theorems, and functional expansions for nonlinear systems. IEEE Centennial Medalist, Former Vice Chairman IEEE Group on Circuit Theory, and Former Guest Editor, IEEE Transactions on Circuit Theory Special Issue on Active and Digital Networks. Fellow and member, IEEE; member, American Association for the Advancement of Science, Eta Kappa Nu, Sigma Xi, Tau Beta Pi, National Academy of Engineering.

AT&T Technical Journal
Vol. 64, No. 8, October 1985
Printed in U.S.A.

Single-Chip Implementation of Feature Measurement for LPC-Based Speech Recognition

J. G. ACKENHUSEN* and Y. H. OH†

(Manuscript received May 14, 1985)

A single-chip implementation of Linear Predictive Coding (LPC)-based feature measurement for speech recognition, called the Feature Extracting Digital Signal Processor (FXDSP), has been developed by programming the AT&T *DSP20*™ programmable Digital Signal Processor (DSP) and has been verified by both numerical simulation and system use. For identical input, the recognition distance between floating point simulation and the DSP implementation was found to be negligibly small when compared with distances for word matches. The feature-measurement technique is identical to that used in numerical simulations of LPC-based isolated- and connected-word recognition using combinations of dynamic time warping, vector quantization, and hidden Markov modeling. As a result, the FXDSP represents a single-chip common building block for real-time implementation of most speech recognition techniques under investigation at AT&T Bell Laboratories. The FXDSP performs eighth-order LPC analysis on speech received from a standard CODEC. In every frame period (15 ms) it produces a feature vector consisting of the log energy, nine amplitude-normalized autocorrelation coefficients, and nine LPC-based test-pattern coefficients. The feature-measurement program requires 1023 locations of the 1024 available in on-chip program ROM, 211 of 256 available RAM locations, and 75 percent of available real time.

I. INTRODUCTION

Most speech recognition work at AT&T Bell Laboratories has been based on a standard form of feature measurement first proposed by

* AT&T Bell Laboratories. † AT&T Bell Laboratories, now with Texas Instruments.

Itakura.[1] Speech recognition features are computed from an eighth-order Linear Predictive Coding (LPC) calculation on 45-ms analysis frames spaced by 15 ms. The autocorrelation method is used, and the speech is of telephone bandwidth (3.3 kHz) and is sampled at 6667 samples per second.

With this front end, numerical simulations have demonstrated successful word recognition algorithms based on dynamic programming for both isolated words[2] and connected words.[3] Real-time hardware that uses this front end for isolated-word recognition has been reported.[4] More recent simulations have used the same front end in recognizers that use vector quantization for isolated-[5] and connected-word recognition[6] and for recognizers using hidden Markov modeling.[7]

Comparative tests of the LPC front end with a variety of filter banks have found the LPC technique to provide superior performance for complex vocabularies over telephone bandwidths.[8]

This paper describes a real-time implementation of this LPC feature-measurement technique that is of single-chip complexity. The implementation uses a programmable signal processor, the AT&T Bell Laboratories Digital Signal Processor (DSP).[9] In this implementation, called the FXDSP (Feature Extracting Digital Signal Processor), the output continuously provides results of LPC analysis of whatever input signal is present with less than one frame (15 ms) of delay.

1.1 Relation to previous work

An implementation of LPC analysis using two DSP chips was previously described by Daugherty.[10] This used an older version of the programmable signal processor known as DSP-1. The DSP-1 operates at one half the speed (5-MHz clock) and has one half the RAM (128 20-bit words) as the *DSP20*™ signal processor used here, but has the same size program memory (1024 16-bit words). Thus, one *DSP20* signal processor is equivalent to two DSP-1 processors in speed and RAM, but is the same as one DSP-1 in program memory.

A major challenge of the work reported here was to reduce the program size by a factor of 2 to attain single-chip implementation. A second challenge was to combine two separate time scales, that of the input (150 μs) and that of the output (15 ms), which had previously been separated by two DSP-1 processors, into a single processor, the *DSP20* signal processor.

A microprocessor-based implementation of an isolated-word recognizer had partitioned the feature-measurement task between a slower general-purpose 16-bit microprocessor performing decision operations and a faster, special-purpose two-board signal processor performing high-speed repetitive arithmetic.[4] This arrangement is similar to the

original simulation environment of a minicomputer and array processor.

An implementation of 10th-order LPC analysis has been developed for the TMS320* signal processor.[11] The TMS320 signal processor uses a sampling rate of 8 kHz, a frame size of 30 ms, and a frame period of 20 ms. This combination of frame size and period results in a frame overlap of 33 percent, where each sample contributes to an average of 1-1/2 frames. The DSP implementation described here uses a frame overlap of 67 percent, and thus requires three frames of computation to be completed on each sample. However, an increase in recognition error rate accompanies the reduction in computation obtained by a reduction in frame overlap, as shown by numerical simulation.[12] In the TMS320 signal processor implementation, the same circuit also performs pattern matching for connected-word recognition.

In addition to realizations based on programmable signal processors, architectures for single-chip LPC feature extractors that use a custom-built processor have been described.[13]

1.2 Organization of paper

In Section II, we examine the equations of LPC feature measurement. Section III describes the DSP chip and the external circuitry required to do the feature measurement. Section IV describes the architecture of the FXDSP program, and Section V presents some details of program implementation. In Section VI, the comparison of the real-time FXDSP calculation with a floating point simulation is described.

II. LPC FEATURE MEASUREMENT

The requirement of LPC is to determine a unique set of predictor coefficients, a_k, $k = 1, 2, \cdots, p$, that minimize the sum of squared differences, E_n, between actual speech samples, $s(n)$, and approximated speech samples, $\tilde{s}(n)$. The approximated speech samples $\tilde{s}(n)$ are formed from a linear combination of speech samples over a short segment of the speech waveform. Thus, the approximate speech samples are given by

$$\tilde{s}(n) = \sum_{k=1}^{p} a_k s(n-k), \tag{1}$$

where $p = 8$ in this analysis. The task of minimizing the prediction error, E_n, is to choose a_k such that

*Trademark of Texas Instruments.

$$E_n = \sum_m e_n^2(m) \tag{2}$$

$$= \sum_m [s_n(m) - \tilde{s}_n(m)]^2 \tag{3}$$

$$= \sum_m [s_n(m) - \sum_{k=1}^{p} a_k s_n(m-k)]^2 \tag{4}$$

is a minimum.

Techniques for calculating the linear prediction coefficients, a_k, from the speech samples, $s(n)$, are described in the literature.[14] The method used here is a block-processing technique based on the autocorrelation method and Durbin's recursion (Fig. 1).

Speech which has been bandlimited to 100 to 3300 Hz and sampled at 6667 samples per second is first preemphasized with a first-order network:

$$s'(n) = s(n) - as(n-1); \quad a = 0.95. \tag{5}$$

The preemphasized speech is then blocked into frames of 300 samples (45 ms) which are spaced by 100 samples (15 ms). Thus, the lth frame of speech, $\tilde{x}_l$, is given by

$$\tilde{x}_l = s'(Ml + n), \; n = 0, 1, \cdots, N-1; \qquad l = 0, 1, \cdots, L-1, \tag{6}$$

where $M = 100$ and $N = 300$ for an input sequence length of L frames. As a result of this choice of M and N, each speech sample contributes to three consecutive analysis frames.

Each frame is then smoothed by a Hamming window:

$$x_l(n) = w(n) \cdot \tilde{x}_l(n), \tag{7}$$

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right), \qquad N = 300. \tag{8}$$

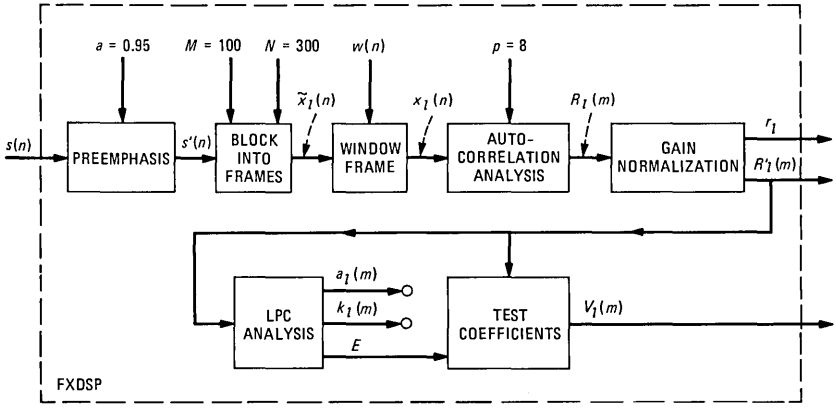


Fig. 1—Signal processing for extracting LPC features for recognition.

The resulting windowed frames of speech data are used to perform an autocorrelation calculation, given by

$$R_l(m) = \sum_{n=0}^{N-1-m} x_l(n)x_l(n+m); \qquad m = 0, 1, \cdots, 8. \tag{9}$$

The logarithm of the frame energy, $R_l(0)$, is then calculated:

$$r_l = \log_2 R_l(0). \tag{10}$$

The autocorrelation coefficients are gain-normalized such that $R'_l(0) = 1$, as follows:

$$R'_l(m) = \frac{R_l(m)}{2^{r_l}}. \tag{11}$$

This normalization is required so that later computation of Durbin's recursion uses the full integer precision of the machine. The log energy, r_l, is used for end-point detection and frame energy information during the recognition process.

Durbin's recursion is then applied to calculate a set of PARCOR coefficients, k_i, $i = 1, 2, \cdots, 8$, and a prediction residual from the $R'_l(m)$ for each frame as follows (the frame index l is suppressed):

$$E^{(0)} = R'(0). \tag{12}$$

For $i = 1, 2, \cdots, 8$, do eqs. (13) through (16):

$$k_i = \frac{\left[R'(i) - \sum_{j=1}^{i-1} \alpha_j^{(i-1)} R'(i-j) \right]}{E^{(i-1)}} \tag{13}$$

$$\alpha_i^{(i)} = k_i \tag{14}$$

$$\alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{i-1}; \qquad (j = 1, 2, \cdots, i-1; \quad i \neq 1) \tag{15}$$

$$E^{(i)} = (1 - k_i^2) E^{(i-1)}. \tag{16}$$

Extract final residual, E, and LPC coefficients a_j:

$$E = E^{(8)} \tag{17}$$

$$a_j = a_j^{(8)}. \tag{18}$$

Test-pattern coefficients are then formed by computing:

$$V_l(m) = \frac{R'_l(m)}{E}, \qquad m = 0, 1, \cdots, 8. \tag{19}$$

The FXDSP output consists of r_l, $R'_l(m)$, and $V_l(m)$ for $m = 0, 1, \cdots, 8$. The PARCOR coefficients k_i and LPC coefficients a_j are calculated as a result of calculating E; however, since they are not used directly in real-time pattern matching, they are discarded. Ref-

erence templates are made up of autocorrelations of a_j that are produced during a non-real-time vocabulary training session. In the current robust training algorithm, a reference pattern is made up of an autocorrelation average of two tokens that correspond to two different repetitions of a word.[15] Therefore, no use can be made of the LPC coefficients in real time.

III. HARDWARE

The hardware for this implementation consists of a μ-law CODEC with filters, which is run at a 6.667-kHz sampling rate, and the AT&T Bell Laboratories DSP, which is run at 10 MHz (Fig. 2). Separate oscillators control the sampling rate of the CODEC and the clock of the DSP.

A design alternative would have been to replace the 8-bit μ-law CODEC with a 12- or 13-bit linear analog-to-digital converter. Although a slight amount of quantization error is introduced by the μ-law conversion of the CODEC followed by the conversion back to 13-bit linear representation in the DSP, this error was seen to be minor. The benefit of the economy of a simple hardware interface between the DSP and the CODEC, the lower cost of the CODEC as compared with a 13-bit linear converter, and the fact that any telephone line input to the CODEC has probably already been subjected to conversions from analog to μ-law digital and back justified the slight degradation of waveform.

A block diagram of the DSP is shown in Fig. 3. The version used here, known as the *DSP20* signal processor, is an improved version of the original signal processor described in Ref. 9 in which both speed and RAM size have been doubled.

The *DSP20* signal processor has a 400-ns instruction cycle time. The processor consists of a read/write memory of 256 20-bit words and a mask-programmable program ROM of 1024 16-bit words. Alter-

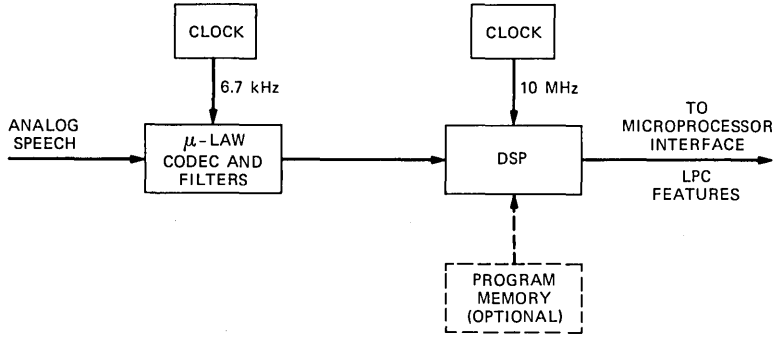


Fig. 2—LPC feature measurement hardware.

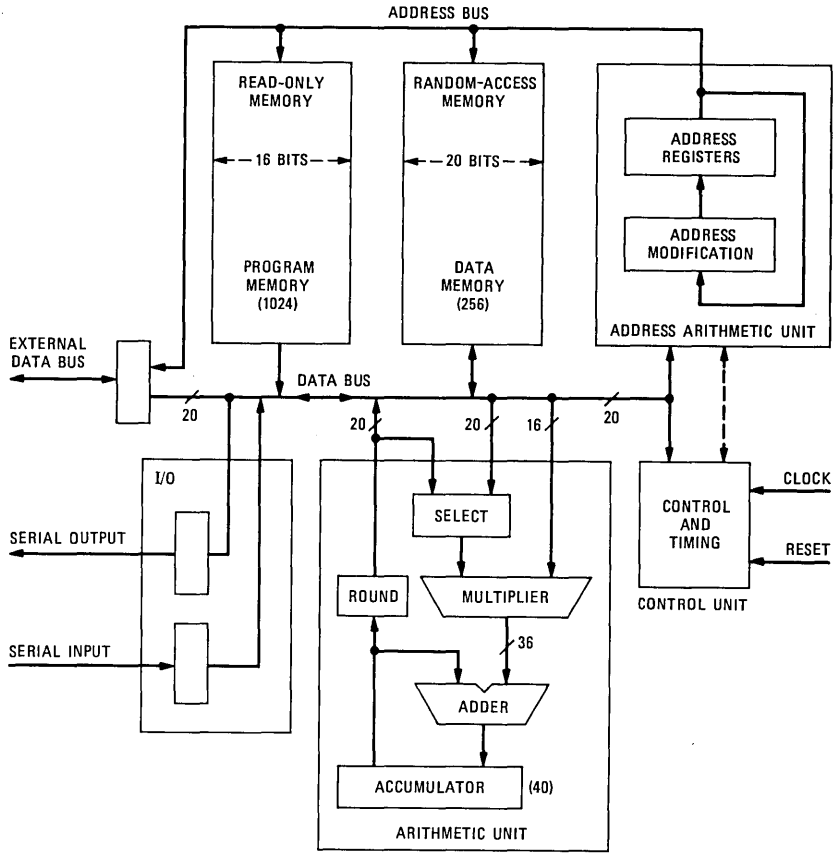


Fig. 3—Block diagram of DSP.

natively, the DSP can be run from 1024 words of external program memory, usually made of erasable programmable ROM, or RAM that can be down loaded. An address arithmetic unit contains registers for controlling memory access. A data arithmetic unit contains a 16-bit × 20-bit multiplier, a 40-bit accumulator, a 40-bit adder, and a 20-bit rounding-overflow circuit. Input and output occur through two serial data pins.

In one 400-ns machine cycle, the DSP can (1) decode an instruction, (2) fetch data and perform a multiplication, (3) accumulate output products from the multiplier, and (4) store data in memory.

IV. PROGRAM ARCHITECTURE

A conflict arises between the input time scale of the FXDSP, one sample every 150 μs, and its output time scale, 19 coefficients of a

feature vector every 15 ms. The FXDSP is required to process a new sample every 150 μs regardless of any other operation in progress, else the input sample is lost and the resulting frame feature vector is incorrect. Thus, two time scales exist, a sample time scale and a frame time scale.

As a result of the two time scales, the program architecture really consists of two separate programs, a sample update program that updates autocorrelation vectors every four samples [eqs. (5) through (9)] and a frame-recursion program that calculates the output feature vector from the autocorrelation vectors from the previous frame [eqs. (10) through (19)]. The frame-recursion program is divided into smaller pieces that are interposed with repeated executions of the sample update program (Fig. 4).

The sample update program operates on four samples each time it is executed. This four-sample operation is a compromise between fully block processing, in which autocorrelation vectors are calculated on a

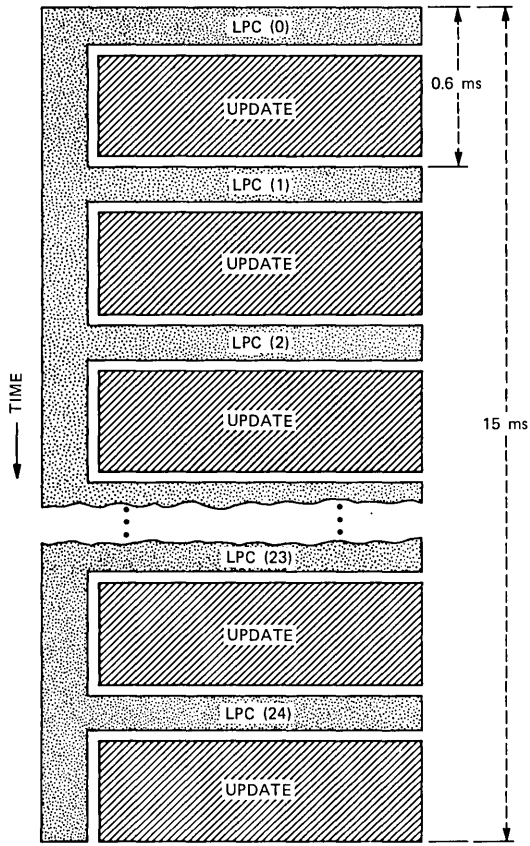


Fig. 4—Interleaving of sample update and frame-inversion programs.

frame of 300 samples all at once, and fully stream processing, in which the autocorrelation vectors are updated upon receipt of each new sample.[10] Fully block processing, however, requires enough read/write memory to store all 300 samples, which is more memory than the single-chip DSP has available. Fully stream processing, which has been used in other implementations,[4,11] executes too slowly for real-time analysis on the DSP.[10] This is because before any autocorrelation update occurs, address pointers must be set up for accessing samples and autocorrelation vectors, and each autocorrelation coefficient must be accessed and placed in the accumulator of the arithmetic unit. These overhead operations are necessary for any number of samples used in the update, and can only be tolerated in real time if the updates occur for more than one sample at a time.

The frame period of 100 samples and the updating of autocorrelation vectors by four samples at a time require that the update program be executed 25 times per frame period. Therefore, an output operation of one frame coefficient is added to the sample update program to provide 25 output coefficients per frame, spaced at four sample intervals. The 19 frame coefficients (r_l, $R_l'(m)$, and $V_l(m)$, $m = 0, 1, \cdots, 8$) and six consecutive zeroes are output for each frame. The sequence of six zeroes provides a synchronization marker for identification of the 19 coefficients by the processor that receives the output of the FXDSP.

Figure 5 shows a more detailed view of the timing of operations. The frame recursion is divided into 25 pieces numbered LPC(0) through LPC(24). Between the first and second samples of the group of four sample inputs, one piece of the frame recursion program is

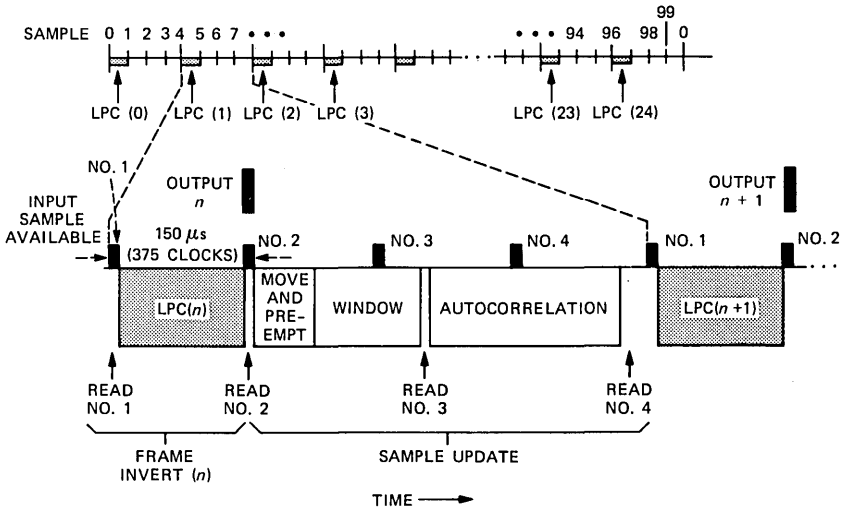


Fig. 5—Timing of input, output, and program sections.

executed. Each piece is completed within the sample period of 150 μs. In Table I, the function of each piece of the frame recursion is shown, as well as the time required for execution and the number of 16-bit words in program ROM required for that piece. As shown at the bottom of Table I, the final 11 time slots, LPC(15) through LPC(24), are unused.

During the time following the second, third, and fourth samples, the sample update operation is performed. The operations associated with the sample update operation are described in Table II. A sample is available every 150 μs and is placed in the FXDSP input buffer by the CODEC. The update program reads that sample at a convenient time, but before the next sample, arriving 150 μs later, overwrites it. Each sample is immediately converted from μ-law to linear encoding by the FXDSP and is then written into a four-sample buffer without any further processing until all four samples are obtained.

Table I—Frame recursion timing and program memory (by function)

Label	Function	Execution Time (μs)	Program Locations
LPC (0)	Read $R_l(m)$ to frame recursion input buffer, shift window	95	97
LPC (1)	Calculate r_l	60	143
LPC (2)	Calculate $R_l'(m)$; $m = 1, 2, 3, 4$	144[1]	50
LPC (3)	Calculate $R_l'(m)$; $m = 5, 6, 7, 8$	144	8[2]
LPC (4)	Set up for Durbin's recursion [$E_0 = R_l'(0)$]	50	32
LPC (5)	Calculate $1/E_{i-1}$ and Durbin's recursion ($i = 1$)	128	226
LPC (6)	Calculate $1/E_{i-1}$ and Durbin's recursion ($i = 2$)	128	6[2]
LPC (7)	Calculate $1/E_{i-1}$ and Durbin's recursion ($i = 3$)	128	6
LPC (8)	Calculate $1/E_{i-1}$ and Durbin's recursion ($i = 4$)	128	6
LPC (9)	Calculate $1/E_{i-1}$ and Durbin's recursion ($i = 5$)	128	6
LPC (10)	Calculate $1/E_{i-1}$ and Durbin's recursion ($i = 6$)	128	6
LPC (11)	Calculate $1/E_{i-1}$ and Durbin's recursion ($i = 7$)	128	6
LPC (12)	Calculate $1/E_{i-1}$ and Durbins's recursion ($i = 8$)	128	6
LPC (13)	Calculate $1/E$	128	6
LPC (14)	Calculate $V_l(m)$, $m = 0, 1, \cdots, 8$	12	41
LPC (15) thru LPC (24)	Idle	12 each	26
	Total (% used of available)	1777 (12%)	688[3] (67%)

[1] For signal 51 dB down from peak; shorter execution time for stronger signals.
[2] Locations include only the subroutine call; subroutine previously counted.
[3] Total includes 17 locations of the power-up initialization routine not listed above.

Table II—Sample update timing and program memory (by function)

Label	Function	Execution Time (μs)	Program Locations
Read #1	Read and μ-to-linear convert sample	3	11
Output	Output one frame feature coefficient	6	29
Read #2	Read and μ-to-linear convert samples		
Move and pre-emp	Shift sample buffer by four samples and preemphasize four samples	60	54
Window	Calculate window values and apply three times to four samples	111	123
Read #3	Read and μ-to-linear convert sample		
Autocorrelation	Use four samples to update nine autocorrelation vectors for three overlapped frames	193	118
Read #4	Read and μ-to-linear convert sample		
	Total (% used of available)	337 (62%)	335 (33%)

As a result, the sample update program has a pipeline delay of four samples. The frame recursion program calculates on the frame just completed and produces the output of a feature vector within one frame period after the end of the corresponding frame.

V. PROGRAM IMPLEMENTATION

This section describes several novel programming techniques that were required to implement the FXDSP. The most scarce resource was program memory; execution time and read/write memory were available in sufficient quantities. Therefore, most innovations were directed toward reducing the amount of program memory required at the expense of increasing execution time or read/write memory requirements. The specifics of program module size, execution time, and execution sequence are covered in Tables I and II.

One major problem, the negotiation between the input sample time scale of 150 μs and the output frame time of 15 ms, was solved by the program architecture discussed in the previous section.

A second problem was the Hamming window computation. Because of the frame size and overlap, each sample falls into the first third of one analysis frame, the second third of the previous analysis frame, and the final third of the twice previous frame. Additionally, after every 100 samples—when one of the three frames is completed—the relationship of the three analysis windows rotates cyclically. As a result, the Hamming window presented both the problem of producing

the cosine-based values and rearranging the segments of the window upon completing a frame.

In earlier implementations,[4,10] the Hamming window was stored as a table in program memory. In this implementation, program memory was too scarce, so a Taylor series expansion was used instead. Each third of the Hamming window (100 samples) was computed from a third-order Taylor series expansion about its midpoint (sample 50, 150, and 250). A comparison of the exact and approximate Hamming window is shown in Fig. 6 in both the time and frequency domain. The approximated window has been slightly shifted up to each comparison—its peak value is actually identical to the peak of the exact window.

To conserve program memory, several pieces of program modules were shared for multiple functions, sometimes with multiple exit points. For example, to perform the division required by eq. (13), the reciprocal of the energy E was calculated. An efficient reciprocal routine developed by Daugherty[16] was used, but required that the number for which the reciprocal was being formed be between 1 and 2. To build a general-purpose reciprocal routine, the number was first normalized to fall within the desired range. The reciprocal was readjusted to its true value to compensate for the normalization. The reciprocal normalization is the same operation as the amplitude nor-

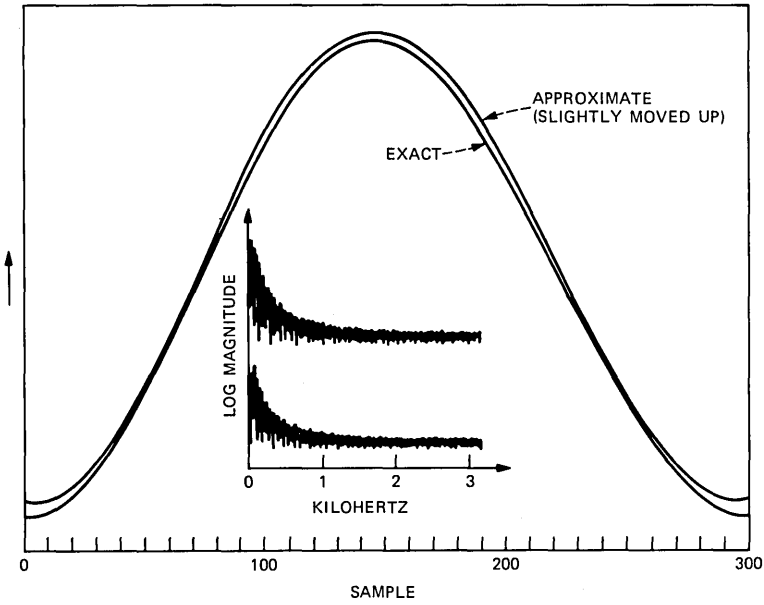


Fig. 6—Comparison of exact and Taylor series approximation of Hamming window.

malization and log energy calculation performed at LPC(1) [eq. (10)], and the same program performs both functions. However, for amplitude normalization, the program is exited before the reciprocal calculation is executed. Thus, the reciprocal routine, which requires 181 program locations, shares 99 of these locations with the gain normalization program, saving on overall program space.

An important way to conserve program space was the development of a means of computing Durbin's recursion with a common piece of code for all orders, $i = 1, 2, \cdots, 8$. Although the recursion is readily executed as a subroutine in microprocessor and Fortran implementations, it is difficult to perform as a subroutine in a programmable signal processor. This is because a DSP does not allow enough addressing capability to handle the two-dimensional array of α and the one-dimensional arrays of k, E, and R. A DSP typically provides only indirect addressing with the ability to increment one of two or three pointer registers by a fixed amount. An implementation of Durbin's recursion, if strung out, requires 536 program locations (not including the reciprocal calculation). With the iteration-independent form used here, that figure drops to 119 program locations.

By careful assignment of memory locations and proper sequencing through the arrays α, k, E, and R, all address calculation was rendered to be sequential within one iteration, that is, only in increments or decrements of one location.[17] This type of address sequencing is within the capability of the signal processor, and makes possible the single subroutine for all iterations. This allowed the frame-recursion program and the sample update program to fit together in the 1024 locations of program memory.

The LPC test coefficients $V_l(m)$ produced by the recursion are scaled by a power of 2 before output to obtain $\tilde{V}_l(m)$:

$$\tilde{V}_l(m) = 2^{-n(m)} \cdot V_l(m), \tag{20}$$

where

$$n(m) = 0; \quad m = 0, 1, 2, 3 \tag{21}$$

$$= 1; \quad m = 4, 5 \tag{22}$$

$$= 2; \quad m = 6 \tag{23}$$

$$= 3; \quad m = 7 \tag{24}$$

$$= 4; \quad m = 8. \tag{25}$$

This scaling is to compensate for a scaling performed on reference coefficients by a factor of $2^{n(m)}$ to allow each reference coefficient to be represented in 12 bits of memory. The values $n(m)$ are based on statistical analysis of the dynamic range of reference coefficients.[12]

To conclude the examination of program implementation, it is important to examine the arithmetic precision used in the signal processing. The μ-law speech is immediately converted to 13-bit linear encoding and is multiplied by 32 to attain an 18-bit word length. All sample update processing before autocorrelation—that is, eqs. (5) through (8)—is performed with 16 bits of precision, with the only approximation being introduced by the Taylor series expansion of the Hamming window. The autocorrelation calculation, eq. (9), is performed with 34 bits of precision, which represents full accuracy for the 13-bit speech samples. Double-precision storage is used on the 34-bit autocorrelation vectors.

A completed frame of autocorrelation vectors is normalized and then truncated to 15 significant bits [eqs. (10) and (11)]. This allows the remaining LPC recursion to be computed on single-precision data. Fifteen-bit precision has been shown to be adequate for fixed-point implementation of Durbin's recursion.[18] The LPC recursion [eqs. (12) through (16)], including the reciprocal calculation, is computed to at least 16 bits of precision. Often, for computations such as the accumulation of sums, eq. (13), the full 40-bit accumulator is used before rounding the sum to the single-word size.

As a result of maintaining full precision throughout the calculation, the difference between the LPC calculation, as computed by the FXDSP and as computed by full-precision floating point simulation, is minimal, as will now be described.

VI. COMPARISON WITH FLOATING POINT SIMULATION

To evaluate the performance of the FXDSP, a comparison of LPC feature measurement as calculated by the real-time FXDSP hardware was compared to LPC feature measurement as calculated by a floating point Fortran simulation running in non-real-time. The input to both routines was a common file of digitized speech, and final comparison was made using the log likelihood spectral distance used in speech recognition. This allowed relative comparison of errors introduced by the FXDSP to typical speech recognition scores.

The two-path program flow is shown in Fig. 7. Input at the left is a linear-encoded, 16-bit-per-sample speech file that had been band-limited to 3.2 kHz and sampled at 6667 samples/s. The program module FORMAT produced two speech files, one in format suitable for down loading into a DSPMATE—a hardware development tool for the AT&T Bell Laboratories DSP—and the other a standard integer speech file for Fortran simulation. Because the FXDSP is intended for use with a μ-law CODEC, one step in the DSPMATE formatting is the conversion of the speech file from linear to μ-law

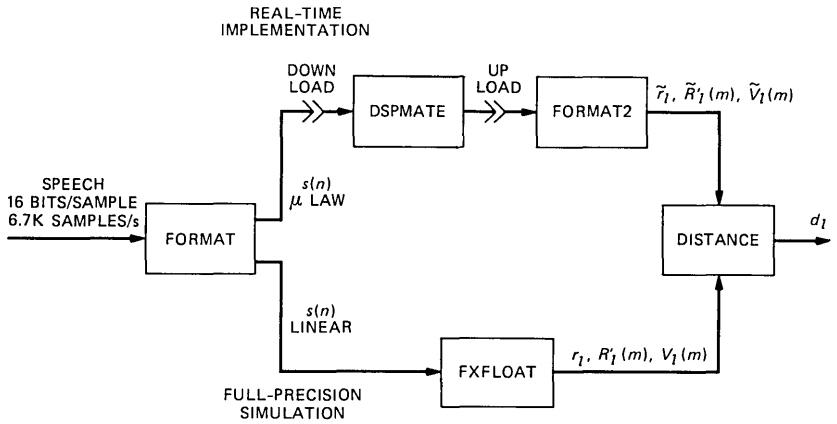


Fig. 7—Program module architecture for FXDSP verification.

encoding. The FXDSP immediately converts each sample back from μ-law to linear.

The upper path represents the route through the DSPMATE. The file capable of being down loaded is sent to the DSPMATE, where it is presented as file input in real time to a DSP chip running the LPC feature-measurement program. The resulting outputs, consisting of log energy $\tilde{r}_l$, gain-normalized autocorrelations $\tilde{R}'_l(m)$, and LPC test-pattern coefficients $\tilde{V}_l(m)$, $m = 0, 1, \cdots, 8$, are then up loaded, reformatted for Fortran simulation (FORMAT2), and input to a log likelihood distance computation program (DIST). The tilde over a quantity indicates that it was calculated by the FXDSP.

The lower route from FORMAT is passed through a floating point computation (FXFLOAT) that produces the values of r_l, $R'_l(m)$, and $V_l(m)$ in a file that is in a format identical to that produced by FORMAT2.

Program DIST computes the log likelihood distance of Itakura[1] for test coefficients produced by the FXDSP and reference coefficients produced by the floating point simulation. Reference coefficients $F_l(m)$ are produced from the LPC coefficients of eq. (18) as follows:

$$F_l(0) = \sum_{j=0}^{8} a_j^2 \tag{26}$$

$$F_l(m) = 2 \cdot 2^{n(m)} \cdot \sum_{j=0}^{8-m} a_j a_{j+m}; \qquad m = 1, 2, \cdots, 8. \tag{27}$$

The values of $n(m)$ are given in eqs. (21) through (25).

The distance calculated by DIST is for test and reference frames taken from the same sequence of speech samples and is given by

$$d_l = \log \sum_{i=0}^{8} F_l(i) \cdot \tilde{V}_l(i). \tag{28}$$

For test and reference coefficients computed with full precision from the same speech samples, $d_l = 0$.

The comparison was performed on 161 frames of speech taken from spoken digits. The dynamic range of the speech was 38 dB.

In Fig. 8a, a histogram of distances computed according to eq. (28) is displayed. The negative distances are a normal result of taking the log of a quantity that is slightly less than 1 due to round-off error. The average of the distances is 0.021.

This distance is negligibly small compared with the distances associated with the variation in word pronunciation shown by scores for correct word recognition. In Fig. 8b, the error histogram of Fig. 8a is overlaid on the histogram for correct word recognition using the same

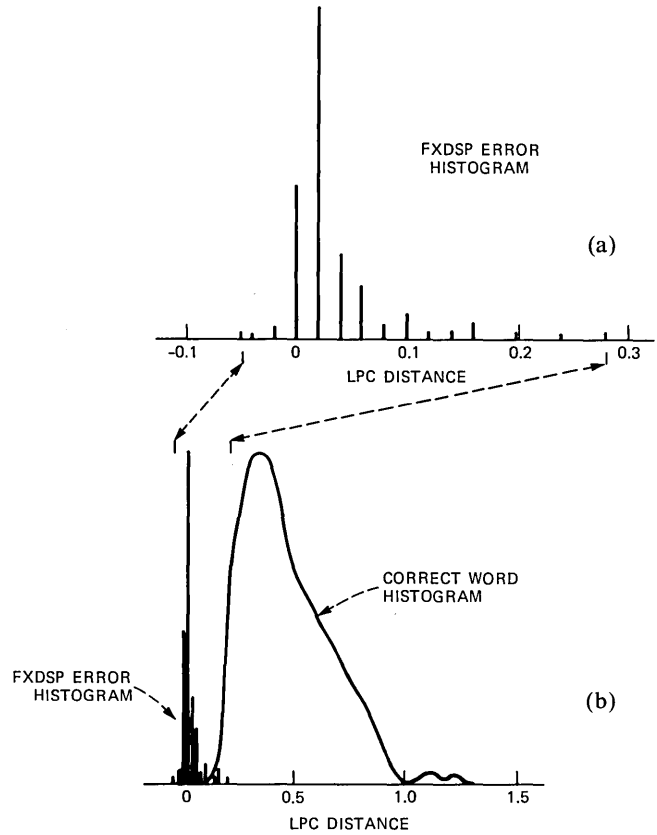


Fig. 8—Comparison of LPC distance from FXDSP to distance of correct word matches.

distance measurement.[19] The average of the correct recognition scores is about 0.45.

The distance of the FXDSP calculation from the true floating point computation is significantly less than the distance arising from variation when a given analog waveform is digitized at randomly varying phase. This distance, measured by repetitively playing taped speech with a professional quality recorder into a digital speech recognizer, averages about 0.035.[20]

The major source of error between floating and FXDSP computation arises from the linear-to-μ-law-to-linear conversion that is performed on the FXDSP path through Fig. 7, but not on the floating point path. Table III shows a sequence of particularly large distances that contributed to Fig. 8 in the left column. In the right column are the much smaller distances that result from performing linear-to-μ-law conversion, followed by μ-law-to-linear conversion, on the speech at a point immediately preceding the floating point LPC analysis (FXFLOAT). The average distance here drops from 0.09 to 0.012. A preliminary investigation on more speech frames suggests that about 75 percent of the distance between floating point simulation and FXDSP implementation is because of the linear-to-μ-to-linear conversion.

VII. SUMMARY

A single-chip basic building block for LPC-based connected- and isolated-word recognition systems has been described. The single chip is an appropriately programmed digital signal processor of AT&T Bell Laboratories.

Because the major limitation in attaining single-chip implementation was the amount of program memory available, several novel programming techniques were used to conserve program memory. These included (1) development of a program architecture that interleaved a background mainframe inversion program with a foreground

Table III—Comparison of FXDSP-to-floating point distances—with and without linear-μ-law-linear conversion in floating point computation

Frame	Without	With
1	0.100	0.013
2	0.147	0.008
3	0.055	0.035
4	0.226	0.008
5	0.052	0.020
6	0.010	0.001
7	0.009	0.002
AVG	0.090	0.012

sample update program, (2) development of a form of Durbin's recursion suitable for implementation as an iteration-independent subroutine, (3) use of overlaid subprograms with multiple exit points, and (4) use of a Taylor series expansion, rather than a look-up table, to store and permute segments of a Hamming window.

Comparison with numerical simulations shows that the error introduced by the implementation is negligible. This good match renders the chip suitable for use in systems that use quantities calculated in floating point on general-purpose computers, such as statistically clustered templates or frames for speaker-independent work recognition or for recognition based on vector quantization or hidden Markov modeling.

REFERENCES

1. F. Itakura, "Minimum Prediction Residual Principle Applied to Speech Recognition," IEEE Trans. Acoust., Speech, Signal Processing, *ASSP-23* (February 1975), pp. 67–72.
2. B. Aldefeld et al., "Automated Directory Listing Retrieval System Based on Isolated Word Recognition," Proc. IEEE, *68*, No. 11 (November 1980), pp. 1364–79.
3. C. S. Myers and L. R. Rabiner, "A Level Building Dynamic Time Warping Algorithm for Connected Word Recognition," IEEE Trans. Acoust., Speech, Signal Processing, *ASSP-29* (April 1981), pp. 284–97.
4. J. G. Ackenhusen and L. R. Rabiner, "Microprocessor Implementation of an LPC-Based Isolated Word Recognizer," Proc. IEEE ICASSP-81 (1981), pp. 746–9.
5. L. R. Rabiner, M. M. Sondhi, and S. E. Levinson, "A Vector Quantizer Incorporating Both LPC Shape and Energy," Proc. IEEE ICASSP-84 (1984), pp. 17.1.1–4.
6. S. C. Glinski, "On the Use of Vector Quantization for Connected-Digit Recognition," AT&T Tech. J., *64*, No. 5 (May–June 1985), pp. 1033–45.
7. L. R. Rabiner, S. E. Levinson, and M. M. Sondhi, "On the Application of Vector Quantization and Hidden Markov Models to Speaker-Independent, Isolated Word Recognition," B.S.T.J., *62*, No. 4 (April 1983), pp. 1075–105.
8. B. A. Dautrich, L. R. Rabiner, and T. B. Martin, "On the Effect of Varying Filterbank Parameters on Isolated Word Recognition," IEEE Trans. Acoust., Speech, Signal Processing, *ASSP-31* (August 1983), pp. 793–807.
9. Special Issue on the Digital Signal Processor, B.S.T.J., *60*, No. 7, Pt. 2 (September 1981), pp. 1431–709.
10. J. W. Daugherty, unpublished work.
11. T. Schalk and M. McMahan, "Firmware-Programmable μC Aids Speech Recognition," Electron. Des., *30* (July 22, 1982), pp. 143–7.
12. L. R. Rabiner, J. G. Wilpon, and J. G. Ackenhusen, "On the Effects of Varying Analysis Parameters on an LPC-Based Isolated Word Recognizer," B.S.T.J., *60*, No. 6 (July–August 1981), pp. 893–911.
13. Y. H. Oh et al., "Architecture for a Real-Time LPC-Based Feature Measurement Integrated Circuit," Proc. IEEE ICASSP-84 (1984), pp. 25B.2.1–4; also B. P. Tao and M. Oijala, "Architecture for a VLSI Implementation of an LPC-Based, Isolated Word Recognition System," Proc. IEEE ICASSP-84 (1984), pp. 34B.5.1–4.
14. L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Englewood Cliffs, N. J.: Prentice Hall, Inc., 1978.
15. L. R. Rabiner and J. G. Wilpon, "A Simplified, Robust Training Procedure for Speaker-Trained, Isolated Word Recognition Systems," J. Acoust. Soc. Amer., *68*, No. 5 (November 1980), pp. 1271–6.
16. J. W. Daugherty, unpublished work.
17. J. G. Ackenhusen, unpublished work.
18. J. D. Markel and A. H. Gray, *Linear Prediction of Speech*, Berlin: Springer-Verlag, 1976.
19. M. K. Brown and L. R. Rabiner, "On the Use of Energy in LPC-Based Recognition of Isolated Words," B.S.T.J., *61*, No. 10 (December 1982), pp. 2971–87.
20. K. L. Shipley, unpublished work.

AUTHORS

John G. Ackenhusen, B.S. (Physics), B.S.E. (Nuclear Engineering), M.S. (Physics), M.S.E. (Nuclear Engineering), 1976; Ph.D. (Nuclear Engineering), 1977, University of Michigan; AT&T Bell Laboratories, 1978—. After serving as Interim Director of the University of Michigan Laser Plasma Interaction Laboratory, Mr. Ackenhusen joined AT&T Bell Laboratories in 1978 and began working in the field of optics for lightwave communications systems. His present activity in computer speech recognition began in 1979 with his interest in real-time hardware for speech recognition, in which he designed the first special-purpose computer for performing speech recognition using the computationally demanding techniques of linear predictive coding and dynamic time warping. In 1981 he became Supervisor of the Speech Recognition Group, where he leads an effort concerned with the development of efficient algorithms, hardware, software, and silicon for real-time speech recognition. Senior Member, IEEE. Member, ASSP Technical Committees on Speech and VLSI, ASSP Conference Board.

Young Hwan Oh, B.S., 1971, M.S., 1974, Ph.D., 1974 (Electrical Engineering), University of New Mexico; M.B.A. Program, 1971–1972; GTE Automatic Electric Laboratories, 1978–1981; AT&T Bell Laboratories, 1981–1984; Texas Instruments, 1984—. While at GTE Automatic Electric Laboratories, Mr. Oh was involved in hardware design, debugging, and testing for an I/O module for the GTD5-EAX, Class 5 Digital End Office Switching System. Also, he was a responsible engineer for converting the analog Dual-Tone Multifrequency (DTMF) receiver to the digital DTMF receiver for No. 5 application. At AT&T Bell Laboratories, he worked in the Speech Recognition Group, where he was involved in speech synthesis and recognition projects. In 1984 he joined Texas Instruments Advanced Technology Laboratory, where he is Director of the Speech Processing Laboratory. His dissertation was in the area of digital filtering and performance analysis. Member, IEEE, Tau Beta Pi.

AT&T Technical Journal
Vol. 64, No. 8, October 1985
Printed in U.S.A.

Blocking When Service Is Required From Several Facilities Simultaneously

By W. WHITT*

(Manuscript received January 14, 1985)

This paper analyzes a mathematical model of a blocking system with simultaneous resource possession. There are several multiserver service facilities without extra waiting space at which several classes of customers arrive in independent Poisson processes. Each customer requests service from one server in each facility in a subset of the service facilities, with the subset depending on the customer class. If service can be provided immediately upon arrival at all required facilities, then service begins and all servers assigned to the customer start and finish together. Otherwise, the attempt is blocked (lost without generating retrials). The problem is to determine the blocking probability for each customer class. An exact expression is available, but it is complicated. Hence, this paper investigates approximation schemes.

I. INTRODUCTION AND SUMMARY

The multifacility blocking problem considered here arises in many contexts and has a long history in traffic engineering (see pages 77 and 95 of Ref. 1). We were motivated by performance analysis issues in packet-switched communication networks. Specifically, we were investigating methods for calculating the blocking probabilities (percentage of failed attempts) in setting up virtual circuits. The need for methods to calculate these blocking probabilities arose in the development of the Packet Network Performance Analysis module of the Packet Network Design and Analysis (PANDA) software package in

* AT&T Bell Laboratories.

the Operations Research Department of AT&T Bell Laboratories.[2,3] It is difficult to analyze the blocking because a circuit typically requires the simultaneous possession of limited resources associated with several different facilities (transmission links, memory buffers, etc.). Moreover, there is competition for the resources not only from other demands for circuits on the same path, but also from demands for different circuits that use only some of the same facilities. Hence, even without alternate routing or waiting (which we do not consider), the blocking is complicated. Our purpose here is to develop bounds and approximations. After we describe our model, we will discuss related work and other applications.

1.1 The mathematical model

There are n multiserver service facilities without extra waiting room and c customer classes. Service facility i has s_i servers. Customers from class j arrive according to a Poisson process with rate λ_j and immediately request service from one server at each facility in a subset A_j of the n service facilities. If all servers are busy in any of the required facilities, the request is blocked (lost without generating retrials). Otherwise, service begins immediately in all the required facilities. All servers working on a given customer from class j start and free up together. The service time for class j at all facilities has a general distribution with finite mean μ_j^{-1}. We assume that the c arrival processes and all the service times are mutually independent.

This model already embodies the extension in which each class requires service from a random subset of the n facilities. Suppose that class j with arrival rate λ_j initially requires one server at each facility in subset A_{jk} with probability p_{jk}, where $\sum_k p_{jk} = 1$. We can represent this more general model within our framework by increasing the number of classes. New class (j, k) has a Poisson arrival process with rate $\lambda_j p_{jk}$ and requires one server in each facility in the subset A_{jk}. This procedure is justified because of two familiar properties of Poisson processes: (1) independent random splitting of a Poisson process produces independent Poisson processes, and (2) the superposition of independent Poisson processes is a Poisson process (see Theorems 4.2 and 5.3 of Ref. 4).

Returning to the previous setting in which each class requires a fixed subset of facilities, we let $b(A)$ be the probability that all servers are busy in at least one facility in the subset A (at an arbitrary time in steady state). Thus $b(i) \equiv b(\{i\})$ is the probability all servers are busy at facility i. Since Poisson arrivals see time averages,[5] $b(A_j)$ is also the blocking probability for class j. [The blocking probability for class j would be $\sum_k p_{jk} b(A_{jk})$ if class j required a random number of facilities, as described above.]

It is not difficult to give the exact formula for $b(A)$ using theory related to queueing networks,[6-8] but the formula is complicated, especially when the numbers n and c are large (see Theorem 4 in Section II). To appreciate the complexity, recall that the arrival rates λ_j, service rates μ_j, and subsets of required facilities A_j for class j can differ from class to class. Hence, our interest centers on developing bounds and approximations.

1.2 Related work

There is a substantial body of related literature. The problem treated here is connected, at least in spirit, to the theory of gradings and link systems in traffic engineering.[1] The specific approximation problem is discussed by Holtzman.[9] Also somewhat related is the work on stochastic models of dynamic storage allocation.[10-12] Previous work also has been done on service systems, with waiting as well as blocking, in which customers require more than one server.[7-20]

Our model is relatively elementary compared with many of these other models. Our analysis benefits by having blocking instead of waiting and by having each customer require exactly one server per facility. On the other hand, we address an issue typically not considered in the papers in which customers require more than one server: Here there are constraints on which servers can be used; there must be one server from each facility. To make the comparison clear, it is useful to modify our model by considering one large facility containing all the servers in all the original facilities. If one of our customers requiring service from one server in each of k facilities could use any k servers in the single large facility, then we would have the model of Arthurs and Stuck.[16] Here, however, there are constraints.

The model considered here is in fact a special case of a more general single-facility blocking model of Kaufman,[7] in which there is a general sharing rule. From Kaufman or Burman, Lehoczky and Lim,[8] we learn that our model is a product-form model with the insensitivity property.[6] This provides expressions for the exact blocking probabilities (Theorem 4 in Section 2.1 here), but as noted above this exact expression is quite intractable. The insensitivity property tells us that the blocking probabilities depend on the service-time distributions only through their means, so that there is no need to assume exponential service-time distributions; for convenience we can replace general service-time distributions by exponential service-time distributions without affecting the blocking probabilities. We discuss insensitivity further in Sections 1.8 and IV.

A special case of our model has also been analyzed in a database locking study by Mitra and Weinberger.[21,22] In their model, the facilities are items in the database and the customers are transactions that

"touch" a specific set of items. To maintain consistency, only one transaction is allowed to touch each item at any time, so that in their model transactions requiring items already being touched are blocked. Their model thus corresponds to the special case of our multifacility blocking model in which each facility has a single server. (It may be of interest to consider the extension of their model to multiserver facilities to represent multiple copies of database items in the database.)

Mitra and Weinberger show that the analysis can be greatly simplified by focusing attention on special symmetric versions of the model. They assume that the arrival rate and service rate for each customer class that requires k facilities (items) is independent of the subset of k facilities required. Moreover, they assume that there is a customer class for each subset of size k. Most important, they consider only the case of one server per facility. (It should be clear that the case of multiserver facilities is much harder.) For these special symmetric models, they obtain an efficient algorithm for calculating the partition function of the product-form model, from which the desired blocking probabilities are easily obtained. [For some database locking applications, it may not be reasonable to assume that the arrival rates for all subsets of size k are identical. Then the approximation methods in this paper may be helpful. See Remark 3 in Section 1.6.]

The symmetric case of the multifacility blocking model has also been considered by Heyman in the investigation of a communication system.[23] We shall also discuss symmetric models here, beginning in Section 1.6. For symmetric models, the approximations are more reliable and much easier to compute.

We have mentioned that this work was primarily motivated by performance analysis issues in packet-switched communication networks, specifically in the PANDA software package.[2,3] The approximations here have also been applied to study the blocking in an AT&T Bell Laboratories computer network[24] and an AT&T Communications model for overseas voice traffic.[25] Another example of the multifacility blocking problem in telephony is contained in Akinpelu.[26] The work that bears most directly on this paper is in Refs. 2, 3, 7, 8, 9, 21, and 23 through 26. (Also see Section VIII.)

1.3 Summary and organization of this paper

We describe our main results in the rest of Section I and provide the supporting technical details in the remaining sections. We discuss three different approximation schemes: the summation bound, the product bound, and the reduced-load approximation. The two bounds are well-known approximations. The reduced-load approximation evidently has a long history,[9] but is not as well known as it deserves to

be. We propose for the reduced-load approximation a successive approximation scheme that is very easy to implement and seems to perform well. In particular, the reduced-load approximation with the successive approximation scheme is ideally suited for large models, where the exact formula becomes intractable.

We obtain two major results about these approximation schemes: (1) As suggested by the names, the first two approximation schemes indeed yield upper bounds on the blocking probabilities, and (2) a limit theorem establishes that the third approximation scheme, the reduced-load approximation, is asymptotically correct for symmetric models as the size of the model grows, in a sense which we will make precise. It is significant that the limiting conditions do not correspond to light traffic as in Refs. 21 and 23, so that in this limit the reduced-load approximation can be very different from the bounds. Our two main results have mathematical interest as well as applied interest, because they are obtained by focusing on multidimensional stochastic processes that are not Markov.

We also obtain two additional light-traffic results. First, we show that all three approximations are asymptotically equivalent as the loads decrease (Corollary 2.3). Second, we show that all three approximations are asymptotically correct as the loads decrease for symmetric models (Corollary 3.2). The qualification "for symmetric models" in the last sentence is important because it can happen for asymmetric models that all three approximations are equally bad in light traffic (see the remark at the end of Section 1.5). As we mentioned in Section 1.2, the approximations are more reliable and easier to use with symmetric models, but we believe they are also very useful for asymmetric models when applied with some caution.

We discuss the bounds in Section 1.4, the reduced-load approximation in Section 1.5, and the main limit theorem in Section 1.6. We discuss numerical examples in Section 1.7; existence, uniqueness, and insensitivity of equilibrium blocking probabilities in Section 1.8; and an extension of the reduced-load approximation for non-Poisson arrival processes in Section 1.9. Additional technical details will be provided in subsequent sections. The main results and directions for future research are summarized in Section VI.

Here are the principal conclusions from our analysis and limited numerical experience: For light loads, for example, blocking in the order of 0.01 or less, the elementary bounds are usually adequate approximations for engineering purposes. Having established that these approximations are bounds, we gain some peace of mind in knowing that the approximations are conservative. For higher levels of blocking, such as 0.05 and above, the reduced-load approximation typically does much better than the elementary bounds. Moreover, the

successive approximation scheme proposed for the reduced-load approximation is very easy to use (Theorem 2), so that the reduced-load approximation seems very attractive when the loads need not be light.

1.4 Bounds

Let $B(s, \alpha)$ be the classical Erlang blocking formula associated with the M/G/s/loss service system with s servers and offered load α,[27,28] defined by

$$B(s, \alpha) = (\alpha^s/s!)/\sum_{k=0}^{s} (\alpha^k/k!), \tag{1}$$

where, as usual, the offered load α is the arrival rate multiplied by the expected service time. Let C_i be the set of all classes that request service from facility i, that is,

$$C_i = \{j : i \in A_j\}. \tag{2}$$

Let $\hat{\alpha}_i$ be the offered load at facility i (not counting blocking elsewhere), defined by

$$\hat{\alpha}_i = \sum_{j \in Ci} \alpha_j, \tag{3}$$

where $\alpha_j = \lambda_j/\mu_j$ is the offered load of class j to the system as a whole.

In Section II we establish the following bounds. These bounds are standard approximations that have long been regarded as conservative.[1] We show that intuition is correct in this case.

Theorem 1: (Product Bound) For each subset A,

$$b(A) \leq 1 - \prod_{i \in A} [1 - B(s_i, \hat{\alpha}_i)].$$

Corollary 1.1: (Facility Bound) For each i, $b(i) \leq B(s_i, \hat{\alpha}_i)$.

Proof: Let $A = \{i\}$. □

Corollary 1.2: (Summation Bound) For each subset A,

$$b(A) \leq \sum_{i \in A} B(s_i, \hat{\alpha}_i).$$

Proof: The summation bound in Corollary 1.2 is always greater than or equal to the product bound in Theorem 1, as is easily verified by induction on the number of facilities in A. Corollary 1.2 also follows directly from Corollary 1.1 and the Bonferroni inequalities (see page 110 of Ref. 29). □

Remarks: For the special case of two facilities, Corollary 1.2 has been proved by different methods by D. D. Sheng and D. R. Smith; see the appendix in Ref. 3. The simple approximation provided by the summation bound in Corollary 1.2 was used in early versions of the PANDA software package,[2] before being replaced by the product bound

in Theorem 1.[3] The summation bound in Corollary 1.2 coincides with the asymptotic approximation developed by Mitra and Weinberger,[21] which is a light-traffic limit. (See Corollaries 2.3 and 3.2 below.)

We give a separate proof of the facility bound in Corollary 1.1, which is of independent interest. We apply Theorem 5 of Smith and Whitt[30] to obtain a monotone-likelihood-ratio ordering for the number of busy servers (Theorem 5), which does not follow from our proof of Theorem 1.

Our proof of Theorem 1 is based on a general technique for comparing a non-Markov process to a Markov process using the conditional transition rates, which applies to many different definitions of stochastic order (Theorem 6). We apply this technique to prove Theorem 1 using the version of stochastic order for probability distributions on R^n based on comparing cumulative distribution functions (Theorem 7). For $n > 1$, this stochastic ordering is weaker than the standard form of stochastic order based on all increasing sets. Our general approach for comparing a non-Markov process to a Markov process has much wider applicability, and is discussed further elsewhere.[31] Our approach exploits stochastic monotonicity of the Markov process,[32–35] and is closely related to the stochastic comparisons for multidimensional Markov processes by Massey that have been applied to establish comparisons for Markovian queueing network models.[36–38]

The bound in Theorem 1 corresponds to independent blocking in the different facilities with the bound Corollary 1.1 used in each facility. It is natural to conjecture that Theorem 1 might be obtained from the more easily established Corollary 1.1 and the inequality

$$b(A) \leq 1 - \prod_{j \in A} [1 - b(i)] \tag{4}$$

but (4) is *not* valid in general (see Example 6 in Section II).

For typical applications in which the bounds are relatively small and customers require only a few facilities, the bounds usually are excellent approximations (see Corollary 3.2), but the following examples demonstrate that the bounds are not always good approximations.

Example 1: Suppose that all n facilities have s servers. Let there be only one customer class, which requests service from all n facilities. Then $\hat{\alpha}_1 = \alpha_1$ and $b(\{1, \cdots, n\}) = b(1) = B(s, \alpha_1) = B(s, \hat{\alpha}_1)$, so that the bound in Corollary 1.1 is tight (an equality), but the bounds in Theorem 1 and Corollary 1.2 can be poor approximations. □

Example 2: Suppose that there are two facilities and two customer classes. Let $s_1 = 10$, $s_2 = 1$, $A_1 = \{1\}$, $A_2 = \{1, 2\}$, $\alpha_1 = 1$, and $\alpha_2 = 100$. Then $B(s_1, \hat{\alpha}_1) = B(10, 101) \approx 1$, but $b(1) \approx 0$, because at most one class 2 customer can be in service at any time. Hence, in this case the bound in Corollary 1.1 is a very bad approximation. □

1.5 A reduced-load approximation

Since the approximations in Theorem 1 and its corollaries are upper bounds, it is natural to look for reduced values that might be better approximations. One way to do this is to reduce the offered load $\hat{\alpha}_i$ at facility i by taking into account the blocking elsewhere. It is natural to develop such an approximation within a framework of facility independence, that is, the assumption that the events of blocking at the difficult facilities are independent. This seems to be a reasonable approximating assumption for "typical" examples, which has been applied before for multiple facilities.[1,14] As a consequence, we have the facility-independence approximation

$$1 - b(A) \approx \prod_{i \in A} [1 - b(i)]. \tag{5}$$

Next we introduce the following approximate total offered load at facility i using the facility independence assumed above:

$$\overline{\alpha}_i = \sum_{j \in C_i} \alpha_j \prod_{\substack{k \in A_j \\ k \neq i}} [1 - b(k)]. \tag{6}$$

In (6) we have reduced the offered load α_j of class j at facility i by the blocking elsewhere. Of course, using (6) we make the offered loads dependent on the blocking probabilities as well as vice versa. [However, the facility-independence approximation in (5) greatly reduces the complexity.] Hence, this leads to a system of equations characterizing the blocking probabilities as our approximation. In particular, our proposed reduced-load approximation for the blocking probability at facility i is the solution to the following system of equations:

$$b^*(i) = B\left\{s_i, \sum_{j \in C_i} \alpha_j \prod_{\substack{k \in A_j \\ k \neq i}} [1 - b^*(k)]\right\}, \qquad 1 \leq i \leq n. \tag{7}$$

From (1), we see that (7) yields n polynomial equations in the n unknowns $b^*(1), \cdots, b^*(n)$.

In general, solving a system of n nonlinear equations in n unknowns can be quite unpleasant. Of course, in many situations there are symmetries in the model, which allow us to reduce the number of equations (and variables). In fact, in the next section we discuss the totally symmetric model, for which (7) reduces to one equation in one variable, which is trivial to solve. (The database model in Ref. 21 also simplifies in this way.) However, we also propose a relatively simple computational scheme for solving the general system (7). In particular, we suggest using successive approximations, that is, iteratively applying the right side of (7) to successive candidate vectors of blocking probabilities. The following theorem indicates that (7) always has a

solution and that the successive approximation scheme either finds the unique solution or provides upper and lower bounds on all solutions to (7).

Theorem 2: (Existence and Successive Approximation) If s_i *and* $\hat{\alpha}_i$ *are strictly positive for each* i*, then the system of eqs.* (7) *has a solution* $\mathbf{b}^* \equiv [b^*(1), \cdots, b^*(n)]$ *with* $0 < b^*(i) < 1$ *for all* i*. All solutions* $\mathbf{b}^*$ *can be bounded above and below, and sometimes found, by successive approximation, that is, by iteratively applying the operator* $T \equiv T\{[b(1), \cdots, b(n)]\}$ *mapping* $[0, 1]^n$ *into itself defined by the right side of* (7)*, starting with* $\mathbf{1} \equiv (1, 1, \cdots, 1)$*. In particular, successive applications of* T *yield the following upper and lower bounds on* $[b^*(1), \cdots, b^*(n)]$ *for all* k*:*

$$(0, 0, \cdots, 0) \equiv \mathbf{0} = T(\mathbf{1}) < T^{2k+1}(\mathbf{1}) < T^{2k+3}(\mathbf{1}) < [b^*(1), \cdots, b^*(n)] < T^{2k+2}(\mathbf{1}) < T^{2k}(\mathbf{1}) < T^0(\mathbf{1}) = \mathbf{1} \equiv (1, 1, \cdots, 1). \quad (8)$$

Proof: First, the operator T defined by the right side of (7) obviously maps $[0, 1]^n$ into itself. Since T is continuous, T has a fixed point, by the classical Brouwer fixed point theorem.[39] Let $\mathbf{b}^*$ represent such a fixed point. Since the operator T is strictly decreasing, $b(i) > b^*(i) > T(\mathbf{b})_i$ for all i, where $\mathbf{b} \equiv (b(1), \cdots, b(n))$, whenever $b(i) > T(\mathbf{b})_i$ for all i and $b(i) < b^*(i) < T(\mathbf{b})_i$ for all i whenever $b(i) < T(\mathbf{b})_i$ for all i. Since $T(\mathbf{1}) = \mathbf{0}$ and $T(\mathbf{0}) < \mathbf{1}$, $0 < b^*(i) < 1$ for all i. □

Since T is continuous and strictly decreasing, the iterated operator $T^{(2)}$ is continuous and strictly increasing. Hence, the successive approximation scheme (8) converges in the sense that $T^{2k+1}(\mathbf{1}) \to \mathbf{L}$ and $T^{2k}(\mathbf{1}) \to \mathbf{U}$, where $\mathbf{L} = (L_1, \cdots, L_n)$ and $\mathbf{U} = (U_1, \cdots, U_n)$ are lower and upper bonds, respectively, on any solution to (7), that is, $L(i) \leq b^*(i) \leq U(i)$, $1 \leq i \leq n$. Often we will have $\mathbf{L} = \mathbf{U}$, that is, $L(i) = U(i) = b^*(i)$ for all $\mathbf{i}$, but not always, because $T^{(2)}$ can have more than one fixed point, as Example 3 below illustrates. Of course, from the monotonicity just discussed, it is clear that the successive approximation scheme in (8) converges if and only if the two-step operator $T^{(2)}$ has only one fixed point.

We have yet to thoroughly investigate when T has a unique fixed point and when the successive approximation scheme (8) converges. Sufficient conditions for T to be a contraction map on a complete metric space—so that T has a unique fixed point and the successive approximation algorithm in (8) converges to it—are given in Section V, but these conditions are very strong. We make the following conjecture. (It has been proved; see Section VIII.)

Conjecture 1: The reduced-load system of eqs. (7) *always has a unique solution.*

It is easy to see that (7) has a unique solution in the case of two

facilities each with a single server. Extensive numerical testing supports Conjecture 1 when there are only two facilities.

It is, of course, natural to wonder whether the model itself might have multiple equilibrium points, but the exact stochastic process under consideration representing the number of customers from each class in service (with exponential service-time distributions) is an irreducible finite-state, continuous-time Markov chain, which necessarily has a unique equilibrium distribution (Section 2.1 below). Thus, if there are multiple solutions to (7), then they must be an artifact of the approximation.

We now present an example to show that the successive approximation in (8) can fail to converge.

Example 3: (Nonconvergence) To see that the succesive approximation scheme in (8) need not converge to a solution of (7), consider the symmetric model with three facilities, each with one server. Let there be only one customer class, which requires service from all facilities. Let the offered load be α. Then (7) consists of the three equations

$$b^*(1) = B\{1, \alpha[1 - b^*(2)][1 - b^*(3)]\}$$
$$b^*(2) = B\{1, \alpha[1 - b^*(1)][1 - b^*(3)]\}$$
$$b^*(3) = B\{1, \alpha[1 - b^*(1)][1 - b^*(2)]\}$$

in the three unknowns $b^*(1)$, $b^*(2)$, and $b^*(3)$. However, when we apply the operator T, we see that T maps the space of vectors (b_1, b_2, b_3) with $b_1 = b_2 = b_3$ into itself. Since we start with (1, 1, 1) in (8), we only need consider the associated operator $\hat{T}$ on [0, 1], defined by

$$\hat{T}(b) = B[1, \alpha(1 - b)^2] = \frac{\alpha(1 - b)^2}{1 + \alpha(1 - b)^2}.$$

The equation $\hat{T}^{(2)}(b) = b$ leads to the polynomial equation

$$x^5 - x^4 + 2\alpha^{-1}x^3 - 2\alpha^{-1}x^2 + (\alpha + 1)\alpha^{-2}x - \alpha^{-2} = 0$$

for $x = 1 - b$, which factors as

$$(x^2 - x + \alpha^{-1})(x^3 + \alpha^{-1}x - \alpha^{-1}) = 0.$$

The second cubic factor also arises as the solution to $\hat{T}(b) = b$. This cubic polynomial is easily seen to be monotone, so that it has a unique root, which falls in the interval (0, 1). This is the unique symmetric fixed point to the symmetric model. The quadratic term has two roots

$$x = (1 \pm \sqrt{1 - 4\alpha^{-1}})/2,$$

which are real and distinct when $\alpha > 4$. These two roots x_1 and x_2 satisfy $0 \leq x_1 \leq 1$ and $x_1 + x_2 = 1$. The quadratic factor does not have real roots when $\alpha < 4$.

In the case $\alpha = 10$, $\hat{T}$ has unique fixed point $b = 0.607$, which corresponds to the symmetric solution (0.607, 0.607, 0.607). However, $\hat{T}^{(2)}$ has three fixed points: 0.113, 0.607, and 0.887. Hence, the successive approximation scheme (8) fails to converge to the unique symmetric fixed point of T; instead it eventually oscillates between $\mathbf{L} =$ (0.113, 0.113, 0.113) and $\mathbf{U} =$ (0.887, 0.887, 0.887). The exact blocking probability in this case is 0.909, obtained directly from the Erlang loss formula (1). The reduced-load approximation for the customer blocking probability is $b^*(\{1, 2, 3\}) = 1 - (1 - 0.607)^3 = 0.939$. □

Remark: It is significant that with exactly two facilities, the successive approximation scheme in (8) converges if and only if T has a unique fixed point, that is, if and only if (7) has a unique solution. We have already noted that convergence of (8) is equivalent to $T^{(2)}$ having only one fixed point. Obviously, $T^{(2)}$ inherits all fixed points of T, so that if T has multiple fixed points, the (8) will not converge. On the other hand, if (8) fails to converge, then the bounds (L_1, L_2) and (U_1, U_2) obtained from (8) are two distinct fixed points of $T^{(2)}$. In turn, (L_1, U_2) and (U_1, L_2) are two distinct fixed points of T.

This argument extends to certain multifacility models, which include many applications of interest.[25] Suppose that the set of facilities can be partitioned into two subsets such that each customer requires service from one facility in each subset. Let there be n_1 facilities in the first subset, numbered $1 \le i \le n_1$, and n_2 facilities in the other subset, numbered $n_1 + 1 \le i \le n_1 + n_2$. If the successive approximation (8) fails to converge, then $(L_1, \cdots, L_{n_1}, L_{n_1+1}, \cdots, L_{n_1+n_2})$ and $(U_1, \cdots, U_{n_1}, U_{n_1+1}, \cdots, U_{n_1+n_2})$ are distinct fixed points of $T^{(2)}$. It is easy to see that $(L_1, \cdots, L_{n_1}, U_{n_1+1}, \cdots, U_{n_1+n_2})$ and $(U_1, \cdots, U_{n_1}, L_{n_1+1}, \cdots, L_{n_1+n_2})$ are then distinct fixed points of T. □

To summarize the proposed reduced-load approximation, we find approximate blocking probabilities at each facility i by solving (7). To solve (7), we suggest using the successive approximation (8). Successive iterations yield upper and lower bounds on all solutions to (7). If the upper and lower bounds are sufficiently close, then we can stop and use the approximation with some confidence. If the successive approximation bounds are not close, then the whole approach is suspect and we suggest using any solution to (7) with caution. An advantage of solving (7) by (8) is that if (8) converges, then we know there is a unique solution to (7). Moreover, if (8) fails to converge, then we get a warning about the whole approach. Also, (8) is extremely easy to implement. Of course, if (8) fails to converge, then we can look for solutions to (7) by other methods. Alternatively, we might choose to use the final upper bound obtained from (8).

After obtaining the approximate blocking probabilities at each facility, [which usually is a solution to (7), but might not be], we obtain

the approximate total offered loads at each facility via (6) and the approximate blocking probabilities for each class via (5).

Remark 1: To implement the successive approximation in (8) or otherwise solve (7), we need to be able to conveniently calculate the Erlang blocking probability eq. (1) and, for some methods such as Newton's method, its derivatives. For this purpose, we can apply techniques of Jagerman.[27,28] □

Remark 2: The successive approximation in (8) and associated bounds closely parallels a proposed successive approximation algorithm to approximately solve closed networks of queues with a decoupling infinite-server node in Section VI of Ref. 40. The analog in Ref. 40 of the operator T above necessarily has a unique fixed point and the successive approximation scheme also yields bounds on it. However, the successive approximation scheme in Ref. 40 also can fail to converge to the fixed point. Further discussion of the successive approximation in Ref. 40 will appear in a subsequent paper. □

Theorem 2 provides a way to relate the reduced-load approximation (7) to the bound in Corollary 1.1. In particular, we can bound the reduced-load approximation (7) much as we already bounded the exact blocking probability at facility i.

Corollary 2.1: The reduced-load blocking approximation at facility i, that is, any $b^*(i)$ obtained from (7), satisfies

$$B\left\{s_i, \sum_{j \in C_i} \alpha_j \prod_{\substack{k \in A_j \\ k \neq i}} [1 - B(s_k, \hat{\alpha}_k)]\right\} < b^*(i) < B(s_i, \hat{\alpha}_i).$$

Proof: The upper bound is $T^2(\mathbf{1})$ and the lower bound is $T^3(\mathbf{1})$ in the successive approximation (8). □

Let $b^*(A)$ be the reduced-load approximate blocking probability for the subset A obtained by combining (5) and (7). From (5) and Corollary 2.1, we immediately obtain the following bounds for $b^*(A)$.

Corollary 2.2: For each subset A, the reduced-load blocking approximation $b^*(A)$ satisfies

$$1 - \prod_{i \in A} \left(1 - B\left\{s_i, \sum_{j \in C_i} \alpha_j \prod_{\substack{k \in A_j \\ k \neq i}} [1 - B(s_k, \hat{\alpha}_k)]\right\}\right)$$

$$\leq b^*(A) \leq 1 - \prod_{i \in A} [1 - B(s_i, \hat{\alpha}_i)].$$

Note that we have not yet given any lower bounds for the exact blocking probabilities. Obviously, $b(A) \geq \max\{b(i): i \in A\}$, but it seems hard to obtain an improvement. One might conjecture that the exact blocking probability $b(i)$ at facility i is bounded below by the lower

bound in Corollary 2.1, but the following example shows that this is not the case.

Example 4: To see that the lower bound in Corollary 2.1 is not a lower bound on the exact blocking probability, suppose that there are three facilities and two customer classes. Let $s_1 = s_2 = 1$, $s_3 = 3$, $A_1 = \{1, 3\}$, $A_2 = \{2, 3\}$, and $\alpha_1 = \alpha_2 = \alpha$. Then $b(3) = 0$ because at most two of the three servers can be busy at the third facility because of the constraints elsewhere. However, it is easy to see that the lower bound in Corollary 2.1 is strictly positive. □

As a further consequence of Theorem 2, we can show that the bounds in Theorem 1 and its corollaries and the reduced-load approximation in (7) are all asymptotically equivalent as the offered loads per facility become negligible, that is, as $\hat{\alpha}_i \to 0$ for all i.

Corollary 2.3: If $\hat{\alpha}_i \to 0$ *for each* i, *then*

$$\text{(i)}\quad B(s_i, \hat{\alpha}_i)/(\hat{\alpha}_i^{s_i}/s_i!) \to 1,$$

$$\text{(ii)}\quad b^*(i)/B(s_i, \hat{\alpha}_i) \to 1,$$

$$\text{(iii)}\quad \left\{1 - \prod_{i \in A} [1 - B(s_i, \hat{\alpha}_i)]\right\} \Big/ \sum_{i \in A} B(s_i, \hat{\alpha}_i) \to 1$$

$$\text{(iv)}\quad b^*(A) \Big/ \sum_{i \in A} b^*(i) \to 1$$

$$\text{(v)}\quad b^*(A) \Big/ \left\{1 - \prod_{i \in A} [1 - B(s_i, \hat{\alpha}_i)]\right\} \to 1$$

for all subsets A, *where* $b^*(i)$ *and* $b^*(A)$ *are the reduced-load approximations based on* (5) *and* (7).

Proof: Part (i) follows immediately from the form of the Erlang blocking formula in (1). Part (ii) follows from Corollary 2.1 after dividing each term by $B(s_i, \hat{\alpha}_i)$ and letting $\hat{\alpha}_i \to 0$ for all i. To establish the limit for the lower bound, let $\bar{\alpha} = \max\{\hat{\alpha}_i,\ 1 \le i \le n\}$ and $\bar{s} = \min\{s_i,\ 1 \le i \le n\}$. Then

$$\prod_{\substack{k \in A_j \\ k \ne i}} [1 - B(s_k, \hat{\alpha}_k)] \ge [1 - B(\bar{s}, \bar{\alpha})]^{n-1}$$

for all i and j, and $[1 - B(\bar{s}, \bar{\alpha})]^{n-1} \to 1$ as $\hat{\alpha}_i \to 0$ for all i. Hence,

$$B\left\{s_i, \sum_{j \in C_i} \alpha_j \prod_{\substack{k \in A_j \\ k \ne i}} [1 - B(s_k, \hat{\alpha}_k)]\right\} \ge B\{s_i, \hat{\alpha}_i[1 - B(\bar{s}, \bar{\alpha})]^{n-1}\}$$

and

$$B\{s_i, \hat{\alpha}_i[1 - B(\bar{s}, \bar{\alpha})^{n-1}]\}/B(s_i, \hat{\alpha}_i) \to 1$$

as $\hat{\alpha}_i \to 0$ because $B(s_i, \hat{\alpha}_i x)/B(s_i, \hat{\alpha}_i) \to x^{s_i}$ as $\hat{\alpha}_i \to 0$ uniformly in x in any compact subinterval of $(0, \infty)$. Given (i) and (ii), (iii) through (v) are elementary. □

Corollary 2.3 demonstrates that using the more elementary approximations in Theorem 1 and its corollaries instead of the reduced-load approximation (7) is justified if the loads are sufficiently light. Theorem 1 and Corollary 2.3 also suggest that the reduced-load approximation $b^*(A)$ itself might be an upper bound, but the following example shows that the reduced-load approximation $b^*(i)$ obtained from (7) is not an upper bound in general.

Example 5: To see that the reduced-load approximation is not an upper bound, let there be two facilities, each with one server. Let there be two classes with $A_1 = \{1, 2\}$ and $A_2 = \{1\}$, so that $\hat{\alpha}_1 = \alpha_1 + \alpha_2$ and $\hat{\alpha}_2 = \alpha_1$. The reduced-load approximation is determined by the two equations

$$b^*(1) = B\{1, \alpha_1[1 - b^*(2)] + \alpha_2\}$$

$$b^*(2) = B\{1, \alpha_1[1 - b^*(1)]\},$$

from which we easily deduce that $0 < b^*(i) < 1$ for each i, so that $b^*(A_2) = b^*(1) < B(1, \alpha_1 + \alpha_2) = b(1)$. □

Remark: One might conjecture that all the approximations for the exact blocking probability $b(i)$ are asymptotically correct as the loads decrease, but Example 2 shows that this is not nearly the case. If $\alpha_2 = 100\alpha_1$ there, then $b(1)/\alpha_1^{10} \to 101$, while $B(s_1, \hat{\alpha}_1)/\alpha_1^{10} \to (101)^{10}$ as $\alpha_1 \to 0$. However, a positive result for large symmetric models appears in Corollary 3.2 below. □

1.6 Symmetric solutions to symmetric models

In this section we consider the special case of symmetric models in which all facilities have s servers and offered load $\hat{\alpha}$, and all classes require service from m facilities. To have full model symmetry, we also assume that there is a class requiring service from each subset of m facilities, and that the offered loads are the same for each class. We also assume that the arrival rates and service rates are the same for all classes.

If we restrict attention to symmetric solutions to symmetric models, then the reduced-load system of eqs. (7) simplifies to the single polynomial equation in one variable

$$b^* = B(s, \hat{\alpha}(1 - b^*)^{m-1}), \tag{9}$$

where $b^*(i) = b^*$ for all i. Since the right side of (9) is continuous and decreasing as a function of b^*, (9) has a unique solution, which is easy to find.

Note that we have restricted attention to symmetric solutions of (7) in order to obtain the single eq. (9). We have not yet ruled out asymmetric solutions to symmetric models. However, we conjecture that none exist. (See Section VIII.)

Conjecture 2 (Corollary to Conjecture 1): The symmetric solution [that is, the solution to (9)*] to the reduced-load approximation eqs.* (7) *is the only solution for a symmetric model.*

To investigate the accuracy of the approximation (5) through (9), we investigate the asymptotic behavior of symmetric models as $n \to \infty$ with the offered load per facility, $\hat{\alpha}$, and the number of facilities required per class, m, held fixed. In Section III we prove that the approximation (5) through (9) is asymptotically correct as $n \to \infty$ under these conditions. Note that since we fix the offered load per facility, $\hat{\alpha}$, this limit does not correspond to light traffic.

To state the main result, let Y_{ni} be the number of busy servers at facility i and let Z_{nk} be the proportion of the facilities with k busy servers (both in steady state) when there are n facilities. Let $\xrightarrow{p}$ denote convergence in probability.

Theorem 3: If $n \to \infty$ with $\hat{\alpha}$ and m held fixed for the symmetric model, then

(a) $Z_{nk} \xrightarrow{p} \beta_k$ as $n \to \infty$ for each k, where β_k satisfies the M/G/s/loss formula

$$\beta_k = (\xi^k/k!) \Big/ \sum_{l=0}^{s} (\xi^l/l!) \tag{10}$$

with

$$\xi = \hat{\alpha}(1 - \beta_s)^{m-1}; \tag{11}$$

that is, β_s is the unique symmetric solution to (9).

(b) For any finite subset H, the random variables Y_{ni}, $i \in H$, are asymptotically mutually independent as $n \to \infty$.

We establish Theorem 3 in Section III by first focusing on the stochastic process representing the proportion of facilities with k busy servers at time t, $1 \le k \le s$ and $t \ge 0$. The key result is a functional law of large numbers for this sequence of stochastic processes as $n \to \infty$ (Theorem 8). The analysis is challenging because this stochastic process is not Markov.

From Theorem 3, we easily obtain our desired corollary.

Corollary 3.1: For symmetric models, the symmetric reduced-load approximation in (5) *through* (9) *is asymptotically correct as $n \to \infty$ with $\hat{\alpha}$ and m held fixed.*

We can combine Corollaries 2.3 and 3.1 to conclude that the bounds in Theorem 1 and its corollaries are also asymptotically correct with light loads.

Corollary 3.2: For symmetric models, the bounds in Theorem 1 *and its corollaries are asymptotically correct as* $n \to \infty$ *and* $\hat{\alpha} \to 0$; *that is, for each integer* k *and each positive* ϵ, *there is a critical offered load* α_0 *and an integer-valued function* $n(\alpha)$ *such that*

$$\left| \frac{b(A)}{kB(s, \hat{\alpha})} - 1 \right| < \epsilon,$$

for all $\hat{\alpha} < \alpha_0$ *and* $n \geq n(\hat{\alpha})$, *where* A *is a subset with* k *facilities.*

Example 5 shows that the reduced-load approximation is not an upper bound on the actual blocking probability in general. Example 1 shows that the symmetric reduced-load approximation $b^*(i)$ for the blocking probability at each facility in a symmetric model need not be an upper bound either. However, we make the following conjecture.

Conjecture 3: The reduced-load approximation for the blocking probability of each customer in a symmetric model in which the number of facilities per customer is fixed, obtained by combining (5) *and* (9), *is always an upper bound.*

Remark 1: In our symmetric model each customer requires service from m facilities. Instead, as in Ref. 21, we could have different types of customers, with customers of type m requiring service from m facilities. The facilities remain symmetric with this change, so that if we still restrict attention to symmetric solutions to the symmetric model, then we again obtain a single polynomial equation in one variable. In particular, suppose that we have M types, numbered from 1 to M. If we let $\bar{\alpha}_m$ be the total offered load of type m at each facility, then we obtain (9) with the second argument of B replaced by $\sum_{m=1}^{M} \bar{\alpha}_m (1 - b^*)^{m-1}$. Consequently, it is easy to approximately solve the models in Ref. 21 and generalizations in which each facility has s servers. With this extended symmetric model we abandon Conjecture 3. It is easy to get a counterexample by modifying Example 1 to introduce additional customers that require service from only one facility and have negligible offered load. □

Remark 2: The reduced-load approximation has the potential of being a powerful and flexible approximation tool if we judiciously control the amount of symmetry. For example, we can obtain a richer class of database locking models by requiring only partial symmetry. Some regions of the database may be requested much more than others. There may also be a tendency for the items requested in a given transaction to cluster together. These general features can be represented by partitioning the database into mutually exclusive subsets and assuming symmetry only within each subset. In addition, we can introduce various types of transactions, as in Remark 1 above. The partial symmetry causes the reduced-load approximation to be a

system of k equations in k unknowns, where k is the number of subsets in the partition. The number of transaction types does not increase the number of equations. Again, the successive approximation (8) can be applied. □

Remark 3: Mitra and Weinberger established Corollary 3.2 for multiple-customer types in the special case of one server per facility via their asymptotic analysis.[21] Heyman also has a different proof of Corollary 3.2 in the special case of one server per facility, assuming that the total offered load in the network is fixed as $n \to \infty$.[23] □

1.7 A few numerical examples

Table I compares the approximations in Theorem 1 and its corollaries with the reduced-load approximation in (5) through (9) for several symmetric models. The various approximations were calculated "by hand" at the terminal using the Erlang blocking formula algorithms of Jagerman[28] (coded by Moshe Segal). The approximations are all independent of the number of facilities, so n is not specified. Based on Theorem 3, the reduced-load approximation in (9) is asymptotically correct for large n. The offered load per facility $\hat{\alpha}$ in (3) is chosen so that the nominal blocking per facility (the bound in Corollary 1.1) has a specified value: 0.10 in the first six cases, 0.02 in the next three cases, and 0.01 in the last three cases.

From Table I (and intuition), it is apparent that the quality of the bounds as approximations is a decreasing function of the number s of servers per facility, the offered load per facility $\hat{\alpha}$, and the number m of facilities per class. In the case of nominal blocking per facility of

Table I—The approximate blocking probability for each customer class in symmetric models: a comparison of the approximation procedures

Servers per Facility s	Offered Load per Facility $\hat{\alpha}$	Facilities per Class m	Summation Bound in Corollary 1.2	Product Bound in Theorem 1	Reduced-Load Approximation (9)
1	0.11111	2	0.200	0.190	0.175
10	7.51	2	0.200	0.190	0.146
50	49.6	2	0.200	0.190	0.126
1	0.11111	3	0.300	0.271	0.234
10	7.51	3	0.300	0.271	0.178
50	49.6	3	0.300	0.271	0.157
1	0.0204	5	0.100	0.096	0.089
10	5.087	5	0.100	0.096	0.072
50	40.27	5	0.100	0.096	0.057
1	0.010101	2	0.0200	0.0199	0.0197
10	4.464	2	0.0200	0.0199	0.0192
50	37.90	2	0.0200	0.0199	0.0180

Table II—Examples of the successive approximations in (8) for the reduced-load approximation with the symmetric model in the first three cases of Table I

Servers per facility s	1	10	50
Offered load per facility $\hat{\alpha}$	0.11111	7.51	49.6
Facilities per class m	2	2	2
Bound in Corollary 1.2	0.200	0.200	0.200
Bound in Theorem 1	0.190	0.190	0.190
Bound in Corollary 1.1	0.100	0.100	0.100
Iteration one	0.089	0.069	0.051
Iteration two	0.0919	0.078	0.074
Iteration three	0.0916	0.076	0.063
Iteration four	0.0917	0.077	0.068
Reduced-load approx. (9)	0.0917	0.076	0.065
Approximate blocking for each class by (5) and (9)	0.175	0.146	0.126

0.01 and only two facilities required per class (the last three cases), the simple summation bound in Corollary 1.2 seems to be adequate. However, the case of $s = 50$ and $m = 5$ produces perhaps a surprisingly large discrepancy between (9) and the bounds.

Table II displays the outcomes of the successive approximations in (8) applied to the first three cases in Table I. The successive iterations describe the blocking per facility, as in (7) and (9). Then (5) is applied to obtain the blocking per class. From Table II it is apparent that about five iterations yields adequate accuracy, that is, getting close enough to the fixed point (9). In these examples the successive approximation scheme in (8) converges to the unique symmetric fixed point of (9).

Table III compares the approximations with exact blocking probabilities for different numbers of facilities in the special case of a symmetric model with $s = 1$ (one server per facility) and ($m = 2$) (two facilities required per class). When $s = 1$, the exact blocking probability is relatively easy to compute because, with exponential service times having mean one (which we can assume without loss of generality by Theorem 4 and Corollary 4.2), the number of customers in service (which is the number of busy servers divided by m) is a birth-and-death process with death rate $\mu(k) = k$ and birth rate

$$\lambda(k) = (n\hat{\alpha}) \left(\frac{(n - km)(n - km - 1) \cdots (n - km - m + 1)}{n(n - 1) \cdots (n - m + 1)} \right). \quad (12)$$

The data in Table III for this special case were obtained from D. P. Heyman (personal communication). This case is consistent with Theorem 3, which establishes that (9) is asymptotically correct as $n \to \infty$. Table III leads us to conjecture that the exact blocking probability for each class is increasing in n in this case. More generally, we make the following conjecture.

Table III—Comparison of approximations with exact blocking probabilities for symmetric models when $s = 1$ (one server per facility) and $m = 2$ (two facilities required per class)

Number of Facilities n	Offered Load per Facility $\hat{\alpha}$	Exact Blocking Probability	Reduced-Load Approximation (9)	Product Bound in Theorem 1	Summation Bound in Corollary 1.2
2	0.010101	0.0100	0.0197	0.0199	0.0200
4		0.0165	0.0197	0.0199	0.0200
8		0.0183	0.0197	0.0199	0.0200
40		0.0195	0.0197	0.0199	0.0200
100		0.0196	0.0197	0.0199	0.0200
2	0.111111	0.100	0.175	0.190	0.200
4		0.154	0.175	0.190	0.200
8		0.168	0.175	0.190	0.200
40		0.1735	0.175	0.190	0.200
100		0.1744	0.175	0.190	0.200
2	1.0000	0.500	0.618	0.750	1.000
4		0.600	0.618	0.750	1.000
8		0.611	0.618	0.750	1.000
40		0.6168	0.618	0.750	1.000
100		0.6176	0.618	0.750	1.000

Conjecture 4: The exact blocking probability for each customer class in a symmetric model is a nondecreasing function of the number n of facilities when the offered load per facility $\hat{\alpha}$ and the number m of facilities per customer are held fixed.

Remark: Conjecture 3 is a corollary to Conjecture 4 and Theorem 3. □

For typical blocking probabilities (0.001 through 0.2), the quality of the approximations appears to be a decreasing function of the offered load per facility (or nominal blocking probability), but this is evidently not true over the full range. The middle four cases in Table III provide greater relative differences than the last four cases, comparing (9) with the exact values.

As our final example in this subsection, we consider a communication network with traffic from several different sources to a common destination, as depicted in Fig. 1. Traffic from each source needs two lines: one line in a facility associated with that source plus one line in a final facility shared by all sources. When there are n sources, there are n customer classes and $n + 1$ facilities. For each i, $1 \leq i \leq n$, class i requires one server from facility i and one server from facility $n + 1$. Note that this example has the special structure mentioned in the remark following Example 3, so that for the reduced-load approximation the successive approximation scheme in (8) converges if and only if the operator T has a unique fixed point.

Tables IV and V give numerical results obtained by J. T. Wittbold[25] for several cases in which n equals 2 and 3, respectively. We display

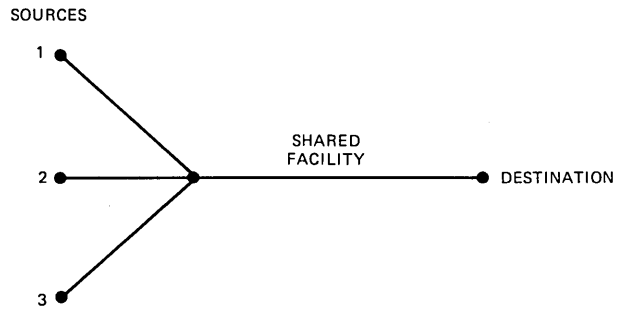


Fig. 1—A communication network with four facilities and three customer classes.

the exact blocking probability and the reduced-load approximation for each facility and for each customer class. The successive approximation converged quickly in every case. For the customer classes, we also display the product bounds and the approximation obtained by taking the product of the exact facility nonblocking probabilities (the last column). This last column helps assess how much of the error is due to assuming facility independence.

For the cases with high blocking probabilities, for example, Case 1 in Tables IV and V, the reduced-load approximation is much better than the product bound, as expected. Overall, the reduced-load approximation appears adequate for engineering purposes. For lighter loads, for example, Cases 4 through 6 in Table IV and V, the product bound seems adequate for most engineering purposes. It should be effective for properly sizing facilities given forecasting data.

1.8 Existence, uniqueness, and insensitivity

It is significant that we have assumed nothing about the service-time distributions except that they have finite means. For applications, experience indicates that call attempts can often be modeled reasonably by a Poisson process, but that virtual circuit holding-time distributions are often not nearly exponential.[41,43] In Section IV we rigorously establish that a steady-state blocking probability exists, is unique, and depends on the service-time distributions only through their means. For this, we apply the theory of Generalized Semi-Markov Processes (GSMPs) and the associated theory of insensitivity.[44–46]

It turns out that the model we consider also can be regarded as a special case of a model analyzed by Kaufman[7] of blocking in a single facility in which customers request several servers and there is a general resource-sharing policy. The connection to Kaufman's single-facility model is made by simply combining our n facilities and implementing our sharing scheme as one of his general sharing policies. The insensitivity property and the exact formula for the blocking are

Table IV—Comparison of approximations with exact blocking probabilities for the communication network example in Fig. 1 with $n = 2$ sources

Case Number	Facility Number	Number of Servers s_i	Offered Load α_i	Blocking Probability at Facility i		Blocking Probability for Customer Class i			
				Reduced Load	Exact	Product Bound	Reduced Load	Exact	Product of Exact Probabilities
1	1	20	30	0.209	0.199	0.61	0.42	0.41	0.43
	2	15	15	0.059	0.025	0.48	0.31	0.30	0.31
	$n+1$	30	—	0.266	0.293	—	—	—	—
2	1	30	30	0.067	0.041	0.29	0.19	0.17	0.19
	2	15	15	0.116	0.101	0.33	0.23	0.23	0.24
	$n+1$	40	—	0.133	0.151	—	—	—	—
3	1	30	30	0.006	0.000	0.45	0.36	0.36	0.36
	2	15	15	0.029	0.017	0.48	0.38	0.37	0.38
	$n+1$	30	—	0.360	0.365	—	—	—	—
4	1	40	30	0.008	0.002	0.068	0.055	0.050	0.051
	2	20	15	0.034	0.030	0.097	0.080	0.075	0.077
	$n+1$	50	—	0.048	0.049	—	—	—	—
5	1	42	30	0.006	0.006	0.023	0.021	0.017	0.017
	2	23	15	0.011	0.012	0.029	0.026	0.023	0.024
	$n+1$	56	—	0.014	0.015	—	—	—	—
6	1	42	27	0.0017	0.0012	0.005	0.005	0.004	0.004
	2	23	13	0.0036	0.0033	0.007	0.007	0.006	0.006
	$n+1$	56	—	0.0027	0.0027	—	—	—	—

Table V—Comparison of approximations with exact blocking probabilities for the communication network example in Fig. 1 with $n = 3$ sources

Case Number	Facility Number	Number of Servers s_i	Offered Load α_i	Blocking Probability at Facility i		Blocking Probability for Customer Class i			
				Reduced Load	Exact	Product Bound	Reduced Load	Exact	Product of Exact Probabilities
1	1	40	30	0.001	0.000	0.30	0.19	0.19	0.19
	2	10	20	0.446	0.452	0.67	0.55	0.55	0.56
	3	20	17	0.027	0.018	0.35	0.21	0.20	0.21
	$n+1$	50	—	0.190	0.192	—	—	—	—
2	1	40	30	0.008	0.003	0.16	0.06	0.05	0.05
	2	10	20	0.523	0.523	0.61	0.54	0.54	0.54
	3	20	17	0.067	0.063	0.22	0.11	0.10	0.11
	$n+1$	61	—	0.044	0.044	—	—	—	—
3	1	40	30	0.012	0.009	0.05	0.03	0.02	0.02
	2	10	8	0.115	0.116	0.15	0.13	0.13	0.13
	3	20	17	0.078	0.079	0.12	0.10	0.09	0.09
	$n+1$	63	—	0.020	0.016	—	—	—	—
4	1	40	30	0.013	0.013	0.03	0.02	0.02	0.02
	2	10	8	0.119	0.121	0.13	0.13	0.12	0.12
	3	20	17	0.083	0.085	0.10	0.09	0.09	0.09
	$n+1$	67	—	0.007	0.003	—	—	—	—
5	1	40	30	0.013	0.011	0.028	0.024	0.019	0.020
	2	10	5	0.017	0.017	0.031	0.028	0.025	0.026
	3	20	13	0.017	0.016	0.031	0.028	0.024	0.025
	$n+1$	60	—	0.011	0.009	—	—	—	—
6	1	40	28	0.0059	0.0045	0.016	0.014	0.011	0.012
	2	10	4	0.0050	0.0048	0.015	0.014	0.012	0.012
	3	20	12	0.0091	0.0084	0.019	0.018	0.016	0.016
	$n+1$	57	—	0.0085	0.0075	—	—	—	—

thus available from Ref. 7. (Reference 7 also mentions other related work.) We contribute to Ref. 7 by verifying the conjecture on p. 1477 there that the insensitivity property holds for arbitrary service-time distributions, not just service-time distributions with rational Laplace-Stieltjes transforms. (The insensitivity analysis for our model extends to the setting of Ref. 7, but the bounds and approximations do not.)

Insensitivity properties in queueing have a long history, going all the way back to Erlang.[47] Insensitivity theory for queueing networks is largely due to Baskett, Chandy, Muntz and Palacios[48] and Kelly.[49] It is now understood[50] that this theory can be viewed as a consequence of the earlier work by Matthes[51] on "bedienungsprozesse" or GSMPs.

As Kaufman observes,[7] his model is equivalent to a closed multiclass BCMP network[48] with the addition of extra population constraints. Without the population constraints, we could simply apply the insensitivity theory developed by Baskett et al.[48] and Kelly,[49] which was extended to arbitrary service-time distributions by Barbour;[52] for example, we could apply Section 3.3 of Ref. 6). As observed by Lam,[53] it is possible to extend the insensitivity theory to closed networks with population constraints, but it is perhaps more appropriate to recognize that the closed network, with or without population constraints, is a GSMP, and the insensitivity theory for GSMPs can be applied directly. The direct approach via GSMPs is contained in Burman et al.[8] The analysis in both Kaufman[7] and Burman et al.[8] requires the addition of Ref. 46 to treat arbitrary service-time distributions. The technical details here for establishing existence, uniqueness, and insensitivity appear in Section IV.

1.9 Non-Poisson arrival processes

We now indicate how the reduced-load approximation (5) through (7) can be combined with previous approximations for the blocking in a single facility with non-Poisson arrival processes to generate approximations for blocking probabilities in the multifacility model here when we relax the assumption that the arrival process of each class is a Poisson process.

We assume that the arrival process of each class is a general stationary point process[54] partially characterized by its arrival rate λ_j and peakedness z_j. (The facilities are thus G/GI/s/loss systems instead of M/GI/s/loss systems. See Refs. 55 through 57 and references in these sources for background on peakedness.) As before, we assume that the arrival processes of the different classes and all the service times are mutually independent.

We regard the arrival process at facility i as the superposition of the arrival processes of those classes requiring service from facility i.

Hence, paralleling (3), we define the peakedness of the arrival process at facility i as

$$\hat{z}_i = \sum_{j \in C_i} \alpha_j z_j / \hat{a}_i, \tag{13}$$

where α_j is the offered load and z_j is the peakedness for class j. Formula (13) is the standard peakedness approximation for a superposition process, but it is based on the assumption that the service rates are the same for all classes, which is not necessarily the case here. Since we do not account for this difficulty, (13) should perform better if the service rates μ_j do not vary much. References 55 through 57 describe ways to determine the peakedness z_j for each class; one relatively simple way is the heavy-traffic approximation in (4) of Ref. 57, but other more involved methods are usually more accurate.

For our new reduced-load approximation, we again use (5) and (6). We propose Hayward's approximation to extend (7).[55-57] However, with the non-Poisson arrival processes we must first carefully distinguish different notions of blocking. Let $b_c(A)$, $b_{Ci}(A)$, and $b_T(A)$ be the probability that all servers are busy in at least one facility in the set A at the instant of an arbitrary arrival, an arrival to facility i, and at an arbitrary time, respectively (the overall call congestion, the facility-i call congestion and the time congestion). Let b_j and $b_j(i)$ be the blocking probability for class j overall and at facility i, respectively. We are primarily interested in b_j, $b_{Ci}(i)$, and $b_T(A)$.

We apply Hayward's approximation to approximate $b_{Ci}(i)$ as if facility i were in isolation. We use the peakedness $\hat{z}_i$ in (13) to modify (7) in the usual way:

$$b_{Ci}(i) = B(s_i/\hat{z}_i,\ \bar{\alpha}_i/\hat{z}_i), \tag{14}$$

where $B(s,\ \alpha)$ is the Erlang blocking formula in (1) extended to noninteger s, as described in Refs. 27 and 28, and instead of (6) $\bar{\alpha}_i$ is

$$\bar{\alpha}_i = \sum_{j \in C_i} \alpha_j \prod_{\substack{k \in A_j \\ k \neq i}} [1 - b_T(k)]. \tag{15}$$

In (15) we use $b_T(k)$ to approximately represent the blocking probability at facility k seen by an arbitrary arrival to facility i. This involves an aspect of the basic facility independence approximation in (5).

We obtain the approximate time congestion for facility i by using the approximation

$$b_T(i) = b_{Ci}(i)/\hat{z}_i \tag{16}$$

(see page 695 of Ref. 57). [A significant improvement should be possible by using (16) of Ref. 56 with the equivalent random method instead of (16) above.] Hence, instead of (7), we obtain the following

system of n equations in the n unknowns $b_{Ci}(i)$ by combining (14) through (16):

$$b_{Ci}(i) = B\left(s_i/\hat{z}_i,\ (1/\hat{z}_i)\sum_{j\in C_i}\alpha_j \prod_{\substack{k\in A_j\\ k\neq i}}\{1 - [b_{Ck}(k)/\hat{z}_k]\}\right). \tag{17}$$

Since $\hat{z}_i$ are fixed positive scalars in (17), the successive approximation in (8) applies here as well; that is, Theorem 2 and Corollaries 2.1 and 2.2 extend easily.

Given that we have obtained $b_{Ci}(i)$ and $b_T(i)$ via (16) and (17), we combine (5) and (16) to obtain the time congestion for an arbitrary subset A, that is,

$$b_T(A) = 1 - \prod_{i\in A}[1 - b_T(i)] = 1 - \prod_{i\in A}\{1 - [b_{Ci}(i)/\hat{z}_i]\}. \tag{18}$$

Next we obtain the blocking for class j at facility i by combining our approximations with Fredericks' approximation for parcel blocking, (23) in Ref. 56. We obtain

$$b_j(i) = b_T(i) + \frac{(z_j - 1)}{(\hat{z}_i - 1)}[b_{Ci}(i) - b_T(i)]. \tag{19}$$

Finally, we obtain the overall blocking for class j by combining (5) and (19), that is,

$$b_j = 1 - \prod_{i\in A_j}[1 - b_j(i)]. \tag{20}$$

The approximations for $b_{Ci}(i)$, $b_T(A)$, and b_j in (17), (18), and (20) have yet to be tested, but experience with the individual approximation steps suggest that the combined procedure is promising.

II. THE BOUNDS

2.1 The exact blocking formula

As a basis for proving Theorem 1, we first calculate the exact blocking probabilities $b(A)$. For this purpose, let N_j represent the steady-state number of class j customers in service. The distribution of the vector $(N_1, \cdots, N_c)$ is conveniently described in terms of the random vector $(N_1^\infty, \cdots, N_c^\infty)$, where N_j^∞ represents the steady-state number of class j customers in service when all n facilities have infinitely many servers, but otherwise the model is the same. Of course, in the infinite-server model the steady-state distribution is easy to describe because there is no blocking, so that there is no interaction among the classes; that is, the random variables $N_1^\infty, \cdots, N_c^\infty$ are independent. From basic results for the M/G/∞ congestion model,[6] the steady-state distribution is

$$P(N_j^\infty = k_j,\ 1 \le j \le c) = \prod_{j=1}^{c} P(N_j^\infty = k_j) = \prod_{j=1}^{c} \left(\frac{\alpha_j^{k_j}}{k_j!}\right). \qquad (21)$$

As in closed Jackson networks of queues, the steady-state distribution of $(N_1, \cdots, N_c)$ is obtained from (21) by simply conditioning. (See Section 1.6 of Ref. 6.) We defer the proof until Section IV.

Theorem 4: The steady-state distribution of $(N_1, \cdots, N_c)$ exists, is unique, depends on the service-time distributions only through their means, and has the form

$$\begin{aligned} &P(N_j = k_j,\ 1 \le j \le c) \\ &= P\left(N_j^\infty = k_j,\ 1 \le j \le c \;\middle|\; \sum_{j \in C_i} N_j^\infty \le s_i,\ 1 \le i \le n\right) \\ &= \frac{P(N_j^\infty = k_j,\ 1 \le j \le c)}{P\left(\sum_{j \in C_i} N_j^\infty \le s_i,\ 1 \le i \le n\right)}. \end{aligned}$$

Of course, Theorem 4 can be used to give an exact expression for the blocking probability $b(A)$. Let Y_i represent the number of busy servers at facility i in our model, that is, $Y_i = \sum_{j \in C_i} N_j$.

Corollary 4.1: For each subset A, $b(A) = 1 - P(Y_i < s_i,\ i \in A)$.

However, we apply Theorem 4 only via the following elementary consequence.

Corollary 4.2: The distribution of $(N_1^\infty, \cdots, N_c^\infty)$ and thus also the distributions of $(N_1, \cdots, N_c)$ and $(Y_1, \cdots, Y_n)$ depend on the vectors of arrival rates $(\lambda_1, \cdots, \lambda_c)$ and service rates $(\mu_1, \cdots, \mu_c)$ only through the vector of offered loads $(\alpha_1, \cdots, \alpha_c)$, where $\alpha_j = \lambda_j/\mu_j$.

Remark: It is significant in Corollary 4.2 that there is not just one degree of freedom, corresponding to the choice of our measuring unit, but c degrees of freedom. For example, we can arbitrarily select the service rate μ_j for each class j, as long as the offered load α_j is as originally specified. In fact, for us it will be convenient to make all service rates identical. (See the proofs of Theorems 5 and 7.)

2.2 Proof of Corollary 1.1

To give a direct proof of Corollary 1.1, we establish a stronger stochastic comparison. Let $N(s, \alpha)$ represent the steady-state number of busy servers in an M/G/s/loss system with s servers and offered load α. We use the notion of Monotone-Likelihood-Ratio (MLR) ordering.[30] An integer-valued random variable X_1 is said to be less than or equal to another integer-valued random variable X_2 in the MLR ordering, denoted by $X_1 \le_r X_2$, if

$$\frac{P(X_1 = k+1)}{P(X_1 = k)} \leq \frac{P(X_2 = k+1)}{P(X_2 = k)} \tag{22}$$

for all k. (We also require that the supports be ordered intervals; that is, $P(X_i = k) > 0$ for integers $k \in [a_i, b_i]$, where $-\infty \leq a_i < b_i \leq +\infty$, $a_1 \leq a_2$, and $b_1 \leq b_2$.) The MLR ordering is useful largely because it implies ordinary *stochastic order*, namely,

$$Ef(X_1) \leq Ef(X_2) \tag{23}$$

for all nondecreasing functions f for which the expectations are well defined.[30,35]

Theorem 5: For each facility i, $Y_i \leq_r N(s_i, \hat{\alpha}_i)$.

Proof: First, make all service-time distributions exponential with mean one, without altering any of the offered loads. By Theorem 4 and Corollary 4.2, this does not alter the steady-state distribution of $(N_1, \cdots, N_c)$. Next apply Theorem 5 of Ref. 30. The service rate in both systems is k when there are k busy servers. The arrival rate at facility i in the actual system is always less than or equal to $\hat{\alpha}_i$. It is less when there is blocking elsewhere. Of course, the support of both random variables is the set $\{0, 1, \cdots, s_i\}$. □

Proof of Corollary 1.1: Apply Theorem 5 and (23), noting that

$$b(i) = P(Y_i \geq s_i) \leq P[N(s_i, \hat{\alpha}_i) \geq s_i] = B(s_i, \hat{\alpha}_i). \quad \square \tag{24}$$

Having proved Corollary 1.1, we immediately obtain Corollary 1.2 by virtue of the Bonferroni inequalities (see page 110 of Ref. 29).

2.3 *Plausible stochastic comparisons*

It is natural to conjecture that Corollary 1.2 could be improved to Theorem 1 by exploiting the exact relationship in Corollary 4.2 and establishing the inequality (4) or, equivalently, that

$$P(Y_i < s_i,\ i \in A) \geq \prod_{i \in A} P(Y_i < s_i). \tag{25}$$

Formula (25) would follow from the random variables Y_i, $i \in A$, being associated or just positively quadrant dependent (see pages 29 and 142 of Ref. 58). Unfortunately, however, (25) is not true in general, as we show in Example 6 below.

One might also try to establish Theorem 1 via certain multivariate stochastic comparisons. In particular, it is natural to consider the multivariate versions of the MLR ordering $\leq_r$ and the stochastic ordering $\leq_{st}$ defined in (22) and (23) (see Refs. 59 and 60). The extension of $\leq_{st}$ is defined again by (23). It is natural to conjecture that

$$(Y_1, \cdots, Y_n) \leq_{st} [N_1(s_1, \hat{\alpha}_1), \cdots, N_n(s_n, \hat{\alpha}_n)], \tag{26}$$

where the variables $N_i(s_i, \hat{\alpha}_i)$ are mutually independent. It is also natural to conjecture the weaker relationships

$$P(Y_i \geq k_i,\ 1 \leq i \leq n) \leq \prod_{i=1}^{n} P[N(s_i, \hat{\alpha}_i) \geq k_i] \tag{27}$$

and

$$P(Y_i \leq k_i,\ 1 \leq i \leq n) \geq \prod_{i=1}^{n} P[N(s_i, \hat{\alpha}_i) \leq k_i] \tag{28}$$

for all n-tuples $(k_1, \cdots, k_n)$. However, in Example 6 below we show that (27) is not valid, which implies that (26) and the stronger ordering with $\leq_r$ instead of $\leq_{st}$ in (26) are not valid either. However, it turns out that (28) is valid, and that is the key to establishing Theorem 1.

Example 6: To see that (25) and (27) need not hold, consider the symmetric model with $n = c = 3$, $s_1 = s_2 = s_3 = 1$, $A_1 = \{1, 2\}$, $A_2 = \{1, 3\}$, $A_3 = \{2, 3\}$, $\mu_1 = \mu_2 = \mu_3 = 1$, and $\lambda_1 = \lambda_2 = \lambda_3 = \alpha$. Then $b(A_j) = 3\alpha/(1 + 3\alpha)$ for all classes j and $b(i) = 2\alpha/(1 + 3\alpha)$ for all facilities i. Hence, for $\alpha > 1$

$$1 - b(A_1) = \frac{1}{1 + 3\alpha} < \left(\frac{1 + \alpha}{1 + 3\alpha}\right)^2 = [1 - b(1)][1 - b(2)], \tag{29}$$

so that (25) fails. On the other hand,

$$1 - b(A_1) = \frac{1}{1 + 3\alpha} > \left(\frac{1}{1 + 2\alpha}\right)^2 = [1 - B(s_1, \hat{\alpha}_1)]^2 \tag{30}$$

so that the conclusion of Theorem 1 still holds in this case.

To see that (27) can fail too, let $(k_1, k_2, k_3) = (1, 1, 0)$. Then

$$P(Y_i \geq k_i, \quad 1 \leq i \leq 3) = \alpha/(1 + 3\alpha), \tag{31}$$

while

$$\prod_{i=1}^{3} P(N(s_i, \hat{\alpha}_i) \geq k_i) = [2\alpha/(1 + 2\alpha)]^2. \tag{32}$$

Hence, for $\alpha^2 < 1/8$, (27) fails. On the other hand, it is easy to see that (28) does still hold in this example. By symmetry, it suffices to consider only the two triples (1, 1, 0) and (1, 0, 0). □

In summary, Example 6 shows that none of the plausible relations (4), (25), (26), and (27) is valid, but the validity of Theorem 1 and (28), which would imply Theorem 1, remains open. We now proceed to establish (28).

2.4 Proof of Theorem 1

To prove Theorem 1 we establish (28). To establish (28), we develop

a certain multivariate variation of Theorem 5 in Ref. 30. In particular, we develop a general stochastic comparison result for continuous-time non-Markov jump processes in which the intensities of moving into certain sets are always greater for one process than the other. The results here are a special case of the general theory developed in Ref. 31. They also can be obtained from the related work of Massey.[36–38]

For our comparison result, we consider an arbitrary finite state space S. (It will be clear that similar results hold for infinite state spaces, but it suffices for us to consider a finite state space.) Let the space $\mathscr{P} \equiv \mathscr{P}(S)$ of all probability measures P on S be endowed with an order relation $\leq$ defined by $P_1 \leq P_2$ if $P_1(A) \leq P_2(A)$ for all subsets A of S in some class $\mathscr{A}$. (The order relation $\leq$ is obviously reflexive and transitive, but it is not necessarily a partial order because it need not be antisymmetric: $P_1 \leq P_2$ and $P_2 \leq P_1$ together do not necessarily imply that $P_1 = P_2$; the relation will be a partial order if $\mathscr{A}$ is a determining class.[61]) Since S is finite, the order relation is closed, that is, it is preserved under limits: If $P_{1n} \leq P_{2n}$ in $(\mathscr{P}, \leq)$ for all n, $P_{in}(\{s\}) \to P_i(\{s\})$ as $n \to \infty$ for each i and $s \in S$, then $P_1 \leq P_2$. [In our application S will be a finite subset of R^n, but $\leq$ will not correspond to ordinary stochastic order on $\mathscr{P}(S)$ as defined in (23).]

The first process $Y_1(t)$ will be a Continuous-Time Markov Chain (CTMC) with infinitesimal transition rates (generator) $q_1(s; A)$, defined as usual for $s \in S$ and $A \subseteq S$ in terms of its transition function by

$$P(Y_1(t + h) \in A \mid Y_1(t) = s) = hq_1(s; A) + o(h), \qquad (33)$$

for $s \notin A$, where $o(h)$ represents a quantity that converges to zero after dividing by h.

The second process $Y_2(t)$ will also be a continuous-time jump process with the jumps governed by infinitesimal transition rates, but as in Ref. 30 these rates may depend on additional information other than the current state, such as the history of the process. Let the additional information at time t be $\Gamma(t)$, and let γ represent a possible value. [In our application the process $Y_2(t)$ represents the number of busy servers at each facility, and the additional information $\Gamma(t)$ is the number of customers of each class in service.] We assume that the process $[Y_2(t), \Gamma(t)]$ is a CTMC on the product state space $S \times S'$, where S' as well as S is finite. Let $q_2(s, \gamma; A)$ be the transition function for $[Y_2(t), \Gamma(t)]$, defined by

$$P([Y_2(t), \Gamma(t)] \in A \mid Y_2(t) = s, \Gamma(t) = \gamma) = hq_2(s, \gamma; A) + o(h) \quad (34)$$

for $(s, \gamma) \notin A$ and $A \subseteq S \times S'$. We shall also use the transition function for $Y_2(t)$, defined by $q_2(s, \gamma; A \times S')$ for $A \subseteq S$ and $s \notin S$.

Just as in Ref. 30, the idea here is to compare the processes $Y_1(t)$

and $Y_2(t)$ by comparing the transition intensities in the space S, requiring that the comparisons hold uniformly in the extra information $\Gamma(t)$, which must be added to $Y_2(t)$ to make $Y_2(t)$ Markov. Of course, a major complication here is the multidimensional state space S. Following Kester[34] and Massey,[36–38] we exploit nonstandard stochastic orderings on S [consistent with (28)] and stochastic monotonicity of the Markov process in this ordering in order to cope with the dimension of the state space. In particular, in our theorem, we shall assume that the transition function of the CTMC $Y_1(t)$ is stochastically monotone.[32–38]

Definition 1: A CTMC $Y_1(t)$ has a stochastically monotone transition function (kernel) $K_t \equiv K_t(s, A) \equiv P(Y_1(t) \in A \mid Y_1(0) = s)$ if $P_1K_t \leq P_2K_t$ in $(\mathscr{P}, \leq)$ whenever $P_1 \leq P_2$ in $(\mathscr{P}, \leq)$, where $(P_iK_t)(A) = \sum_{s\in S} P_i(s)K_t(s, A)$.

Remark 1: It is significant in Definition 1 that both the condition and the conclusion involve the same (unspecified) order relation $\leq$ on $\mathscr{P}$. □

Remark 2: As in Section 2 of Keilson and Kester,[33] stochastic monotonicity of a CTMC $Y_1(t)$ can be characterized by the transition rate function $q_1(s; A)$ and, after uniformization, by the transition function $I + \epsilon q_1$ of an associated discrete-time Markov chain with the same stationary distribution, where I is the identity map and ϵ is sufficiently small so that $I + \epsilon q_1$ is nonnegative. In particular, (i) $(P_1q_1)(A) \leq (P_2q_1)(A)$ for all $A \in \mathscr{A}$ whenever $P_1 \leq P_2$ and (ii) $P_1(I + \epsilon q_1) \leq P_2(I + \epsilon q_1)$ whenever $P_1 \leq P_2$ are each necessary and sufficient for $Y_1(t)$ to have a stochastically monotone transition function. □

For $A \subseteq S$, let $A^c = S - A$. Let $\hat{\pi}_2$ be the marginal distribution of π_2 on S, that is, $\hat{\pi}_2(A) = \pi_2(A \times S')$.

Theorem 6: Suppose that the CTMCs $Y_1(t)$ and $(Y_2(t), \Gamma(t))$ defined above have unique stationary distributions π_1 on S and π_2 on $S \times S'$. If (i) $Y_1(t)$ has a stochastically monotone transition function in $(\mathscr{P}, \leq)$ and (ii) for all $A \in \mathscr{A}$ and $\gamma \in S'$, $q_2(s, \gamma; A \times S') \leq q_1(s; A)$ for all $s \in A^c$ and $q_2(s, \gamma; A^c \times S') \geq q_1(s; A^c)$ for all $s \in A$, then $\hat{\pi}_2 \leq \pi_1$ in $(\mathscr{P}, \leq)$.

Proof: Since π_2 is the unique stationary distribution of $[Y_2(t), \Gamma(t)]$,

$$0 = (\pi_2q_2)(A) = \sum_{s,\gamma} \pi_2(s, \gamma)q_2(s, \gamma; A)$$

for all $A \subseteq S \times S'$. By condition (ii),

$$0 = (\pi_2q_2)(A \times S') \leq \sum_{s,\gamma} \pi_2(s, \gamma)q_1(s, A) = (\hat{\pi}_2q_1)(A) \qquad (35)$$

for all $A \in \mathscr{A}$. Since the transition function associated with q_1 is stochastically monotone, (35) implies that $\hat{\pi} \leq \pi_1$. To see this, let P_{d1}

be the stochastically monotone transition function $I + \epsilon q_1$ of the associated discrete-time Markov chain constructed by uniformization. Then $0 \leq (\hat{\pi}_2 q_1)(A)$ for all $A \in \mathscr{A}$ is equivalent to $\hat{\pi}_2 \leq \hat{\pi}_2 P_{d1}$. Since P_{d1} is stochastically monotone, $\hat{\pi}_2 \leq \hat{\pi}_2 P_{d1} \leq \hat{\pi}_2 P_{d1}^2 \leq \cdots \leq \hat{\pi}_2 P_{d1}^n$. Since $\hat{\pi}_2 P_{d1}^n \to \pi_1$ as $n \to \infty$ and $\leq$ is a closed order, $\hat{\pi}_2 \leq \hat{\pi}_2 P_{d1}^n \leq \pi_1$. □

Remark 1: If $Y_2(t)$ is a Markov processes, so that we do not need $\Gamma(t)$, then Theorem 6 follows from Section 4.2 of Stoyan.[35] In fact, as explained in Ref. 31, Theorem 6 can also be viewed as a consequence of both Stoyan[35] and Massey.[38] □

Remark 2: For both Markov and non-Markov processes, the conditions of Theorem 6 also imply stochastic comparisons for the marginal distributions at time t for all t.[31] □

Remark 3: To relate Theorem 6 here to Theorem 5 of Ref. 30, note that it suffices to let one of the processes there, say $Y_1(t)$, have transition rates that do not depend on the extra information; that is, let $\lambda_1(k, I_t) = \alpha_1(k)$ and $\mu_1(k, I_t) = \beta_1(k)$. (The more general case follows by just making two comparisons.) Then $Y_1(t)$ becomes a birth-and-death process on the integers, which is known to be stochastically monotone with the usual stochastic order for probability measures on the real line. Theorem 6 here thus yields stochastic order (which is weaker than the MLR ordering in Ref. 30) under the conditions of the corollary to Theorem 5 in Ref. 30. Since the stationary distribution of $Y_1(t)$ depends on $\alpha_1(k)$ and $\beta_1(k+1)$ only through the rations $\alpha_1(k)/\beta_1(k+1)$, we can generalize the conditions here to the conditions of Theorem 5 in Ref. 30. In conclusion, then, Theorem 6 here yields a weaker conclusion (stochastic order instead of MLR order) under the same conditions as Theorem 5 of Ref. 30, but Theorem 6 here extends conveniently to the multivariate setting.

We now apply Theorem 6 to our problem. Theorem 1 follows immediately from (28), which we now establish.

Theorem 7: For each n-tuple $\mathbf{k} \equiv (k_1, \cdots, k_n)$, $P(Y_i \leq k_i, 1 \leq i \leq n) \geq \prod_{i=1}^n P[N(s_i, \hat{\alpha}_i) \leq k_i]$.

Proof: We apply Theorem 6. The left and right sides of the inequality will be the stationary distributions of the processes $Y_2(t)$ and $Y_1(t)$, respectively, representing the number of busy servers at each facility for $1 \leq i \leq n$. In both cases, we assume that the service-time distributions are exponential, which we can do without loss of generality by Theorem 4. The process $Y_2(t)$ represents the process of interest to us and the process $Y_1(t)$ is a CTMC in which the coordinate stochastic processes are independent. In other words, $Y_1(t)$ is the process corresponding to n independent M/M/s/loss facilities. The information $\Gamma(t)$ associated with the process $Y_2(t)$ in Theorem 6 here is the number

of class j customers in service for each j at time t. It is easy to see that the process $\Gamma(t)$ and the bivariate process $[Y_2(t), \Gamma(t)]$ are CTMCs.

To fill in the rest of the details, let the state space S be the product of n integer intervals and let the state space S' for $\Gamma(t)$ be the product of c integer intervals, that is,

$$S = \underset{i=1}{\overset{n}{X}} \{0, 1, \cdots, s_i\} \quad \text{and} \quad S' = \underset{i=1}{\overset{c}{X}} \{0, 1, \cdots, \bar{s}\}, \tag{36}$$

where $\bar{s} = \max\{s_i, 1 \le i \le n\}$. Let S be endowed with the usual partial order in R^n; that is, $\mathbf{k}_1 \le \mathbf{k}_2$ for $\mathbf{k}_i = (k_{i1}, \cdots, k_{in})$ if $k_{1j} \le k_{2j}$ for all j. We shall be interested in lower subsets of S defined by

$$L(\mathbf{k}) = \{\mathbf{k}' \in S : \mathbf{k}' \le \mathbf{k}\}. \tag{37}$$

Let $\mathscr{A}$ be the set of complements of lower sets $L(\mathbf{k})$ for $\mathbf{k} \in S$; that is, $\mathscr{A} = \{L(\mathbf{k})^c \equiv S - L(\mathbf{k}) : \mathbf{k} \in S\}$. The set $\mathscr{A}$ induces a partial-order relation $\le$ on the space $\mathscr{P} \equiv \mathscr{P}(S)$ of all probability measures on S through the definition

$$P_1 \le P_2 \quad \text{if} \quad P_1(A) \le P_2(A) \quad \text{for all} \quad A \in \mathscr{A}. \tag{38}$$

Here $\le$ is a proper partial-order relation because $\mathscr{A}$ is a determining class.

It remains to show that conditions (i) and (ii) in Theorem 6 hold with respect to the ordering $\le$ in $\mathscr{P}(S)$. To see that condition (i) holds, that is, that q_1 is stochastically monotone with respect to $\le$, construct the associated discrete-time transition function $P_{d1} = I + \epsilon q_1$ (see Remark 2 before Theorem 6) and note that

$$(\pi P_{d1})[L(\mathbf{k})] = \sum_i p_i^{\pm} \pi[L(\mathbf{k} \pm \mathbf{e}_i)] + \left(1 - \sum_i p_i^{\pm}\right) \pi[L(\mathbf{k})], \tag{39}$$

where $\mathbf{e}_i$ is an n-tuple of all 0's except a 1 in one place and $p_i^{\pm}$ is a probability. (The permissible values of $\pm\mathbf{e}_i$ obviously depend on $\mathbf{k}$, but it is not necessary to specify them or the probabilities $p_i^{\pm}$ in detail.) From (39), it is immediate that $(\pi_1 P_{d1})[L(\mathbf{k})] \ge (\pi_2 P_{d1})[L(\mathbf{k})]$ for all $\mathbf{k} \in S$ if $\pi_1[L(\mathbf{k})] \ge \pi_2[L(\mathbf{k})]$ for all $\mathbf{k} \in S$.

To establish condition (ii), involving the comparison of the intensities, first apply Corollary 4.2 to make all the individual service rates identical without changing the stationary distributions being compared, as in the proof of Theorem 5. Next consider transitions upwards due to arrivals. Observe that for $\mathbf{k} \in L(\mathbf{k}')$

$$q_1[\mathbf{k}; L(\mathbf{k}')^c] = q_2[\mathbf{k}, \gamma; L(\mathbf{k}')^c] = 0 \tag{40}$$

unless $k_i = k_i'$ for some i and

$$q_2[\mathbf{k}, \gamma; L(\mathbf{k}')^c] \le q_1[\mathbf{k}; L(\mathbf{k}')^c] \tag{41}$$

otherwise. Make the comparison (41) by matching the intensities associated with each class j separately. For q_2 this corresponds to a simultaneous jump up of one in all coordinates of A_j with intensity λ_j, while for q_1 this corresponds to a jump up of one in one of the coordinates of A_j, each with intensity λ_j. Strict inequality occurs in (41) if the simultaneous transitions are blocked by the upper boundary, while the corresponding individual transition is not. Inequality also occurs if $k_i = k_i'$ for two or more indices i. Assuming that $k_i = k_i'$ for some i and there is no blocking at the upper boundary, the intensity of transition out of $L(\mathbf{k}')$ is λ_j for q_2 but $m\lambda_j$ for q_1, where m is the number of indices for which $k_i = k_i'$.

Next consider transitions downwards due to departures, where now all individual service rates are identical, say μ. (Invoke Corollary 4.2.) The transition function q_2 differs from q_1 by having multiple departures at intensity μ (that depend on the classes present) instead of individual departures each at intensity μ. The overall intensity of a transition downward, therefore, can be much greater in q_1, but with q_1 it is possible to enter the sets $L(\mathbf{k}')$ from outside, that is, from $\mathbf{k} \in L(\mathbf{k}')^c$ only by a departure in at most one of the coordinates. In other words, we have

$$q_2[\mathbf{k}, \gamma; L(\mathbf{k}')] \geq q_1[\mathbf{k}; L(\mathbf{k}')] \tag{42}$$

for all $\mathbf{k} \in L(\mathbf{k}')^c$. Strict inequality can occur in (42) if $k_i = k_i' + 1$ for two or more i in A_j and $k_i \leq k_i'$ otherwise when a class j customer is in service at time t. Then

$$q_2[\mathbf{k}, \gamma; L(\mathbf{k}')] = \mu > 0 = q_1[\mathbf{k}; L(\mathbf{k}')]. \tag{43}$$

for $\mathbf{k} \in L(\mathbf{k}')^c$. Properties (40) through (43) establish condition (ii) of Theorem 6 in our case. □

III. LARGE SYMMETRIC MODELS

To support the reduced-load approximation in Sections 1.5 and 1.6, we investigate large symmetric models. The limit theorems here are similar in spirit to previous ones for closed networks of queues with unlimited waiting space in Sections V and VIII in Ref. 40.

Here we assume that all facilities have s servers, all service-time distributions are exponential, all service rates are 1, all customer class arrival rates are λ, and all customers require service from m facilities. We associate one class with each possible subset of size m. We let the number of facilities n become large with the total offered load per facility $\hat{\alpha}$ held fixed. We achieve this by letting the arrival rate per class when there are n facilities be

$$\lambda_n = \hat{\alpha} n \Big/ \binom{n}{m} = \frac{\hat{\alpha} m!(n-m)!}{(n-1)!}. \tag{44}$$

It seems useful to focus on the stochastic process $Q_{nj}(t)$ representing the number of facilities with j busy servers at time t in the model with n facilities. Obviously, $Q_{n0}(t) = n - [Q_{n1}(t) + \cdots + Q_{ns}(t)]$ so that it suffices to focus on j with $1 \le j \le s$. The process $[Q_{n1}(t), \cdots, Q_{ns}(t)]$ is convenient because its dimension does not change as $n \to \infty$. It also appears that this process contains the essential information to characterize the asymptotic behavior of the blocking probability. However, this process presents a serious difficulty because, except in the relatively elementary special case in which $s = 1$, this process is not Markov. The future evolution of the process given any present value depends on additional information, namely, the specific classes present. However, we show that in a sense this information is asymptotically irrelevant.

3.1 A conjectured diffusion process limit

Let $V_{nj}(t)$ be the normalized stochastic process defined by

$$V_{nj}(t) = (Q_{nj}(t) - n\beta_j)/\sqrt{n}, \qquad t \ge 0, \tag{45}$$

and let $\mathbf{V}_n \equiv \mathbf{V}_n(t)$ be the vector-valued process defined by

$$\mathbf{V}_n(t) = [V_{n1}(t), \cdots, V_{ns}(t)], \qquad t \ge 0. \tag{46}$$

In the spirit of many limit theorems for closely related Markov processes,[61-63] we conjecture that $\mathbf{V}_n$ converges in distribution to a multivariate diffusion process. It should be possible to establish weak convergence (convergence in distribution) in the function space $D[0, \infty)$ of right-continuous functions with left limits,[61,64,65] but we support the diffusion approximation only by establishing convergence of the infinitesimal means. For the following conjecture, let $\mathbf{V}_n(t)$ be the stationary version (starting in equilibrium at $t = 0$) for each n, which exists and is unique by Theorem 4. The conjectured limit process is an s-dimensional multivariate Ornstein-Uhlenbeck diffusion process, which is characterized by its infinitesimal means and covariances.[62,63,66] The infinitesimal means and covariances have the relatively simple form of $M\mathbf{v}$ and Σ, where $\mathbf{v}$ is the s-dimensional state vector and M and Σ are $s \times s$ matrices that do not depend on the state.

Conjecture 5: The sequence of stationary stochastic process $\{\mathbf{V}_n, n \ge 1\}$ *defined in* (45) *and* (46) *converges weakly (in distribution) in the function space* $D([0, \infty), R^s)$ *to a stationary multivariate Ornstein-Uhlenbeck diffusion process if the normalization constants* β_j *in* (45) *are defined by* (10) *and* (11).

Heuristic Argument: In support of Conjecture 5, we prove that the infinitesimal means of $\{\mathbf{V}_n\}$ converge as $n \to \infty$ to those of an s-dimensional Ornstein-Uhlenbeck diffusion process. Even though the

process $\mathbf{V}_n(t)$ is not Markov for each n, the infinitesimal means depend on the past $\{\mathbf{V}_n(s), s \le t\}$ only through the present state $\mathbf{V}_n(t) = \mathbf{v}$ for each n. For $1 \le j \le s-1$, the infinitesimal means are

$$\begin{aligned}
&\mathbf{m}_{nj}(v_1, \cdots, v_s) \\
&\equiv \lim_{s \to 0} E\left[\frac{V_{nj}(t+s) - V_{nj}(t)}{s} \middle| \mathbf{V}_n(u), u \le t, \mathbf{V}_n(t) = \mathbf{v} \equiv (v_1, \cdots, v_s)\right] \\
&\approx n^{-1/2}\left\{(n\beta_{j-1} + \sqrt{n}v_{j-1})(m\alpha)\left(\frac{n - n\beta_s - \sqrt{n}v_s}{n}\right)^{m-1}\right. \\
&\quad + m(j+1)(n\beta_{j+1} + \sqrt{n}v_{j+1}) - (n\beta_j + \sqrt{n}v_j)(m\alpha) \\
&\quad \left.\cdot\left(\frac{n - n\beta_s - \sqrt{n}v_s}{n}\right)^{m-1} - mj(n\beta_j + \sqrt{n}v_j)\right\} \\
&\approx n^{1/2}\{\beta_{j-1}m\alpha(1-\beta_s)^{m-1} + m(j+1)\beta_{j+1} \\
&\quad - \beta_j m\alpha(1-\beta_s)^{m-1} - mj\beta_j\} + \{v_{j-1}m\alpha(1-\beta^s)^{m-1} \\
&\quad + v_{j+1}m(j+1) - v_j m\alpha(1-\beta_s)^{m-1} - v_j mj\}, \qquad (47)
\end{aligned}$$

where $\beta_0 = 1 - (\beta_1 + \cdots + \beta_s)$. For $j = s$, the infinitesimal mean is

$$\begin{aligned}
&\mathbf{m}_{ns}(v_1, \cdots, v_s) \\
&\approx n^{-1/2}\left\{(n\beta_{s-1} + \sqrt{n}v_{s-1})(m\alpha)\left(\frac{n - n\beta_s - \sqrt{n}v_s}{n}\right)^{m-1}\right. \\
&\quad \left. - ms(n\beta_s + \sqrt{n}v_s)\right\} \\
&\approx n^{1/2}\{\beta_{s-1}m\alpha(1-\beta_s)^{m-1} - ms\beta_s\} + \{v_{s-1}m\alpha(1-\beta_s)^{m-1} - v_s ms\}. \qquad (48)
\end{aligned}$$

In order for $\mathbf{m}_{nj}(v_1, \cdots, v_s)$ to converge as $n \to \infty$, it is necessary and sufficient to have the coefficients of $n^{1/2}$ vanish in the first terms of (47) and (48); that is, we need

$$\begin{aligned}
&\beta_{j-1}\alpha(1-\beta_s)^{m-1} + (j+1)\beta_{j+1} = \beta_j\alpha(1-\beta_s)^{m-1} + j\beta_j, \; j \le s-1, \\
&\qquad \beta_{s-1}\alpha(1-\beta_s)^{m-1} = s\beta_s. \qquad (49)
\end{aligned}$$

By induction, it follows that (10) and (11) provide the unique solution to (49). The remaining terms in (47) and (48) provide the infinitesimal means of the limiting diffusion process.

A next step to establish Conjecture 5 would be to establish convergence of the infinitesimal covariances, but the infinitesimal covariances do depend on more than the current state $\mathbf{v}$ for each n, and seem difficult to calculate. Finally, this would not actually complete the proof because the process $\mathbf{V}_n(t)$ is not Markov. [It almost would if $\mathbf{V}_n(t)$ were Markov by page 268 of Stroock and Varadhan.[63]]

Conjecture 6 (Corollary to Conjecture 5): The stationary random vector of $\mathbf{V}_n(t)$ is asymptotically normally distributed with zero mean vector as $n \to \infty$.

3.2 A law of large numbers

To establish Theorem 3 in Section 1.4, we prove a weaker result than Conjecture 5, namely, a functional law of large numbers for the process $\{[Q_{n1}(t), \cdots, Q_{ns}(t)], t \geq 0\}$ as $n \to \infty$. For this purpose, let

$$X_{nj}(t) = n^{-1}Q_{nj}(t), \qquad 1 \leq j \leq s, \tag{50}$$

and

$$\mathbf{X}_n(t) = [X_{n1}(t), \cdots, X_{ns}(t)] \tag{51}$$

for $t \geq 0$. Note that the components of $\mathbf{X}_n(t)$ are always nonnegative and their sum is at most one, so we can let the state space for $\mathbf{X}_n(t)$ be the s-dimensional simplex, say Δ, which is a compact subset of R^s.

The limiting stochastic process $\mathbf{X}(t)$ for $\mathbf{X}_n(t)$ will be a continuous deterministic motion, that is, a Markov diffusion process with zero diffusion or variance coefficient. The process $\{\mathbf{X}(t), t \geq 0\}$ has a transition function

$$P[\mathbf{X}(t) = T(t, \mathbf{x}) \mid \mathbf{X}(0) = \mathbf{x}] = 1,$$

where $\mathbf{x} \in \Delta$ and $T(t, \cdot)$ is a deterministic function mapping Δ into itself. Let $T_j(t, \mathbf{x})$ be the jth component of $T(t, \mathbf{x})$, that is, $T(t, \mathbf{x}) = [T_1(t, \mathbf{x}), \cdots, T_s(t, \mathbf{x})]$. The function $T(t, \cdot)$ is characterized by its derivative with respect to t, say $T'(\mathbf{x}) = [T_1'(\mathbf{x}), \cdots, T_s'(\mathbf{x})]$, where $T_j'(\mathbf{x}) = d/(dt)T_j(t, \mathbf{x})$, which is independent of t and is essentially the infinitesimal generator. Let $\Rightarrow$ denote weak convergence (convergence in distribution) of random elements in any space, for example, the state space Δ or the space of all sample paths $D([0, \infty), \Delta)$.[61,64,65]

Theorem 8: Assume exponentially distributed service times with mean one. If $\mathbf{X}_n(0) \Rightarrow \mathbf{X}(0)$ in Δ, then $\mathbf{X}_n \Rightarrow \mathbf{X}$ in $D([0, \infty), \Delta)$, where $\mathbf{X}(t)$ is a continuous deterministic motion with transition function $T(t, \mathbf{x})$ having derivatives with respect to t

$$T_j'(\mathbf{x}) = m[\hat{\alpha}(1 - x_s)^{m-1}(x_{j-1} - x_j) + (j + 1)x_{j+1} - jx_j], \; j \leq s - 1,$$

$$T_s'(\mathbf{x}) = m[\hat{\alpha}(1 - x_s)^{m-1}x_{s-1} - sx_s], \tag{52}$$

where $\mathbf{x} = (x_1, \cdots, x_s)$ and $x_0 = 1 - (x_1 + \cdots + x_s)$.

Proof: There are two steps, which we establish in Lemmas 1 and 2 below. First, we show that $\{\mathbf{X}_n\}$ is uniformly tight in $D([0, \infty), \Delta)$, so that every subsequence has a weakly convergent subsequence (see page

35 of Ref. 61). In the process, we show that each limit process has continuous sample paths. Then we show that the transition functions $P[\mathbf{X}_n(t_1 + t_2) \in A \mid X_n(t_1) = \mathbf{x}]$ converge to the transition function of the specified continuous deterministic motion as $n \to \infty$. Moreover, we show that the transition probability is asymptotically Markov, that is, asymptotically independent of the history of the process before t_1.

By Lemma 1, there is a weakly convergent subsequence, and any weakly convergent subsequence, say $\{\mathbf{X}_{n_k}\}$, has some limit process $\mathbf{X}'$. As a consequence of the weak convergence in the function space and the continuous mapping theorem (Theorem 5.1 of Billingsley[61]), the bivariate joint distributions converge weakly in Δ^2 too; that is,

$$P\{[\mathbf{X}_{n_k}(t_1), \mathbf{X}_{n_k}(t_2)] \in \cdot\} \Rightarrow P\{[\mathbf{X}'(t_1), \mathbf{X}'(t_2)] \in \cdot\}$$

for all $t_1, t_2 \geq 0$. Since $\mathbf{X}_n(0) \Rightarrow \mathbf{X}(0)$, $\mathbf{X}'(0)$ must be distributed the same as $\mathbf{X}(0)$. Moreover, since the transition functions converge, the limit $\mathbf{X}'(t)$ must be distributed as $T[t, \mathbf{X}(0)]$. Since the sample paths of $\mathbf{X}'$ are continuous, this determines the distribution of $\mathbf{X}'$ in $D([0, \infty), \Delta)$. Since the distribution of the limit of every weakly convergent subsequence of $\{\mathbf{X}_n\}$ in $D([0, \infty), \Delta)$ is determined, the entire sequence thus converges weakly to the determined limit, by Theorem 2.3 of Billingsley.[61] □

Lemma 1: The sequence $\{\mathbf{X}_n\}$ *is uniformly tight in* $D([0, \infty), \Delta)$ *and the limit of any convergent subsequence has continuous paths.*

Proof: To establish uniform tightness in $D([0, \infty), \Delta)$, we establish the stronger C-tightness, conditions for which are given in Theorem 8.3 of Billingsley.[61] This implies that $\{\mathbf{X}_n\}$ is also D-tight and that the limit of any convergent subsequence has continuous sample paths. To establish C-tightness, it suffices to focus on a single coordinate of $\{\mathbf{X}_n\}$ in $D([0, \infty), R)$, say $\{X_{nj}\}$ (see Section 2 of Ref. 65 and Exercise 6, page 41, of Ref. 61). Moreover, it suffices to restrict the time interval to a compact subinterval.[64,65,67] Since the state space Δ of $\mathbf{X}_n$ is a compact subset of R^s, the set of all probability measures on Δ with the topology of weak convergence is metrizable as a compact metric space (see page 45 of Ref. 68). By Prohorov's theorem, page 37 of Ref. 61, $\{X_{nj}(0)\}$ is uniformly tight in R and condition (i) of Theorem 8.3 in Billingsley[61] holds.

We establish the remaining condition (ii) by bounding the change in $X_{nj}(t)$ in a fixed interval of length δ by the normalized sum of all arrivals and all departures during that arrival. The arrivals, in turn, are bounded by the total number of arrivals that would occur if all servers remained empty throughout the interval, that is, by a Poisson random variable with rate $nm\hat{\alpha}\delta$. Similarly, the number of departures is bounded above by the number of departures that would occur if all facilities remained full throughout the interval, that is, by a Poisson

random variable with rate $nms\delta$. These two bounds can be expressed via stochastic order relations, as in (23), by actually generating the arrivals and departures by appropriately thinning two independent Poisson processes with the indicated rates.[69]

To establish condition (*ii*), it remains to show that for all positive c, ϵ, and η there exists δ such that

$$P[N(cn\delta) > n\epsilon] < \delta\eta \tag{53}$$

for all n sufficiently large, where $N(\lambda)$ is a Poisson random variable with mean λ. Of course, we choose δ so that $c\delta < \epsilon$ to have the mean of $N(cn\delta)$ less than $\eta\epsilon$. Then, using Chebyschev's inequality, we obtain

$$\begin{aligned} P[N(cn\delta) > n\epsilon] &< \frac{\text{Var } N(cn\delta)}{[EN(cn\delta) - n\epsilon]^2} \\ &< \frac{cn\delta}{(cn\delta - n\epsilon)^2} = \frac{c\delta}{n(c\delta - \epsilon)^2}, \end{aligned}$$

which shows that (53) indeed holds for all $n > n_0$, where $n_0 = c/[\eta(c\delta - \epsilon)]$. □

Let A^ϵ be the open ϵ-ball in Δ about the set A, that is,

$$A^\epsilon = \{\mathbf{x} \in \Delta\colon d(\mathbf{x}, \mathbf{y}) < \epsilon \quad \text{for some} \quad \mathbf{y} \in A\}, \tag{54}$$

where d is a metric on R^s, here taken to be the maximum metric $d(\mathbf{x}, \mathbf{y}) = \max\{|x_i - y_i|, 1 \le i \le s\}$.

Lemma 2: For all positive ϵ, states $\mathbf{x} \in \Delta$ and histories $\{\mathbf{X}_n(u), u \le t_1\}$,

$$\lim_{n\to\infty} P(\mathbf{X}_n(t_1 + t_2) \in [T(t_2, \mathbf{x})]^\epsilon \mid \mathbf{X}_n(u), u \le t_1, \mathbf{X}_n(t_1) = \mathbf{x}) = 1,$$

where T is the continuous deterministic motion in Theorem 8.

Proof: Let I be the identity map on Δ. Since $T(t, \cdot)$ has the semigroup property of a Markov process and the derivative T' is bounded and continuous, $(I + \epsilon T')^{t/\epsilon} \to T(t, \cdot)$ as $\epsilon \to 0$. Consequently, it suffices to prove that there is a constant K such that for all sufficiently small positive ϵ

$$\lim_{n\to\infty} P(\mathbf{X}_n(t + \epsilon) \in (\mathbf{x} + \epsilon T')^{K\epsilon^2} \mid \mathbf{X}_n(u), u \le t, \mathbf{X}_n(t) = \mathbf{x}) = 1. \tag{55}$$

To establish (55), we use stochastic dominance arguments as in the proof of Lemma 1. In particular, we first observe that, for any n, t and ϵ, the total number of arrivals in the interval $[t, t + \epsilon]$ is stochastically dominated by a Poisson variable with mean $nm\hat{\alpha}\epsilon$. Similarly, for any n, t, and ϵ, the total number of departures in the interval $[t, t + \epsilon]$ is stochastically dominated by a Poisson variable with mean $nms\epsilon$. These stochastic bounds give us initial bounds on how much $\mathbf{X}_n(u)$ can differ from $\mathbf{X}_n(t)$ in the interval $[t, t + \epsilon]$ for all possible histories. Since

$X_{nj}(t)$ is a proportion, we can apply a law of large numbers for Poisson variables as the rate increases. In particular, there is a constant K, which is independent of ϵ as $\epsilon \to 0$, such that

$$\lim_{n\to\infty} P\Bigg(\sup_{t\le u\le t+\epsilon} |X_{nj}(u) - X_{nj}(t)| > K\epsilon \,|\, \mathbf{X}_n(u), \quad u \le t, \quad \mathbf{X}_n(t) = \mathbf{x} \Bigg) = 0. \quad (56)$$

We now use the initial bound in (56) to produce better bounds on $\mathbf{X}_n(t+\epsilon) - \mathbf{X}_n(t)$, that is, to establish (55). Given that $\mathbf{X}_n(t) = \mathbf{x}$ and

$$\sup_{t\le u\le t+\epsilon} |X_{nj}(u) - X_{nj}(t)| < K\epsilon$$

for $1 \le j \le s$, the actual flow rate into state j (the rate of increase of $X_{nj}(u)$) in the interval $[t, t+\epsilon]$ is bounded above by

$$\begin{aligned} I^u(j) &= \hat{\alpha} \min\{1, (1 - x_s + K\epsilon)^{m-1}\}(x_{j-1} + K\epsilon) \\ &\quad + (j+1)(x_{j+1} + K\epsilon) \\ &\le \hat{\alpha} \min\{1, (1 - x_s + K\epsilon)^{m-1}\}x_{j-1} + \hat{\alpha}K\epsilon + (j+1)x_{j+1} \\ &\quad + (j+1)K\epsilon \\ &\le \hat{\alpha}(1 - x_s)^{m-1}x_{j-1} + (j+1)x_{j+1} + (\hat{\alpha}m + (j+1))K\epsilon \end{aligned} \quad (57)$$

and bounded below by

$$\begin{aligned} I^l(j) &= \hat{\alpha} \max\{0, (1 - x_s - K\epsilon)^{m-1}\}(x_{j-1} - K\epsilon) \\ &\quad + (j+1)(x_{j+1} - K\epsilon) \\ &\ge \hat{\alpha}(1 - x_s)^{m-1}x_{j-1} + (j+1)x_{j+1} - [\hat{\alpha}m + (j+1)]K\epsilon. \end{aligned} \quad (58)$$

In other words, with n facilities the flow into state j for the unnormalized process $Q_{nj}(t)$ is stochastically bounded above by a Poisson process with rate $nI^u(j)$ and stochastically bounded below by a Poisson process with rate $nI^l(j)$.[69] Similarly, the flow rate out of state j [the rate of decrease of $X_{nj}(u)$] in the interval $[t, t+\epsilon]$ is bounded above by

$$\begin{aligned} O^u(j) &= \hat{\alpha} \min\{1, (1 - x_s + K\epsilon)^{m-1}\}(x_j + K\epsilon) + j(x_j + K\epsilon) \\ &\le \hat{\alpha}(1 - x_s)^{m-1}x_j + jx_j + (\hat{\alpha}m + j)K\epsilon \end{aligned} \quad (59)$$

and bounded below by

$$\begin{aligned} O^l(j) &= \hat{\alpha} \max\{0, (1 - x_s - K\epsilon)^{m-1}\}(x_j - K\epsilon) + j(x_j - K\epsilon) \\ &\ge \hat{\alpha}(1 - x_s)^{m-1}x_j + jx_j - (\hat{\alpha}m + j)K\epsilon. \end{aligned} \quad (60)$$

We invoke a well-known functional law of large numbers for the

Poisson process (which is a consequence of the functional central limit theorem, Section 17 of Billingsley[61]) to deduce that as $n \to \infty$ the change in $X_n(u)$, that is, the change in the proportions, is bounded above and below by the deterministic motions with rates $I^u(j) - O^l(j)$ and $I^l(j) - O^u(j)$, respectively. Hence, for any history $\{\mathbf{X}_n(u), u \le t\}$ and any state $\mathbf{X}_n(t) = \mathbf{x}$,

$$\lim_{n\to\infty} P\{\epsilon[I^l(j) - O^u(j)] \le X_{nj}(t+\epsilon) - X_{nj}(t) \le \epsilon[I^u(j) - O^l(j)] \mid \mathbf{X}_n(u), u \le t, \mathbf{X}_n(t) = \mathbf{x}\} = 1, \quad (61)$$

but

$$\epsilon[I^u(j) - O^l(j)] = \epsilon T_j'(\mathbf{x}) + \epsilon^2 K'$$

and

$$\epsilon[I^l(j) - O^u(j)] = \epsilon T_j'(\mathbf{x}) - \epsilon^2 K'$$

for $K' = (2\hat{\alpha}m + 2j + 1)K$, so that (61) is equivalent to the desired result. □

We now describe the limiting continuous deterministic motion $\mathbf{X}(t)$ specified in Theorem 8. In particular, we verify that $\mathbf{X}(t)$ has a unique stationary distribution and converges to it as $t \to \infty$ for any initial distribution. It is relatively elementary that $T(t, \cdot)$ has a unique fixed point in Δ. We want to establish the stronger result that $T(t, \cdot)$ has a unique fixed point in the space $\mathscr{P}(\Delta)$ of all probability measures on Δ. To appreciate the difference, note that clockwise circular motion at constant angular velocity in the plane has a unique fixed point in the plane, namely, the origin, but the uniform distribution over any circle centered about the origin is a stationary distribution for this clockwise circular motion. We show that our continuous deterministic motion actually converges to its unique fixed point in Δ for every initial distribution.

Theorem 9: For any initial vector $\mathbf{y}$, $\mathbf{X}(t) \to \beta$ *as* $t \to \infty$, *where* $\beta \equiv (\beta_1, \cdots, \beta_s)$ *is determined by* (10) *and* (11).

Corollary 9.1: The limiting continuous deterministic motion $\mathbf{X}(t)$ *has a unique stationary distribution, which is a unit mass on the vector* β *determined by* (10) *and* (11).

Proof: We write $T_t(A_1) \to A_2$ as $t \to \infty$ for subsets A_1 and A_2 of Δ to represent that $T(t, \mathbf{y}) \to A_2$ as $t \to \infty$ for all $\mathbf{y} \in A_1$, that is, $d(T(t, \mathbf{y}), A_2) \to 0$ as $t \to \infty$, where $d(\mathbf{x}, A) = \inf\{d(\mathbf{x}, \mathbf{y}): \mathbf{y} \in A\}$ with d the metric on R^{s-1}. Equivalently, $T(t, \mathbf{y}) \to A_2$ as $t \to \infty$ if the limits of all convergent subsequences $\{T(t_k, \mathbf{y}), k = 1, 2, \cdots\}$ of $\{T(t, \mathbf{y}), t \ge 0\}$ with $t_k \to \infty$ are contained in A_2. (Since Δ is a compact metric space, every sequence has a convergent subsequence. Moreover, the limit sets

A_2 considered below will be closed, so that they will contain the limits.) Our goal is to show that $T_t(\Delta) \to \{\beta\}$. To do so, we construct compact subsets $L_1, \cdots, L_s$ such that

$$L_s = \{\beta\} \subseteq L_{s-1} \subseteq \cdots \subseteq L_1 \subseteq \Delta, \tag{62}$$

$T_t(\Delta) \to L_1$ and $T_t(L_k) \to L_{k+1}$, $1 \leq k \leq s-1$, as $t \to \infty$. Since T_t has the semigroup property $T(t_1 + t_2, \mathbf{x}) = T[t_1, T(t_2, \mathbf{x})]$ for all $\mathbf{x}$, t_1 and t_2, and is continuous, this implies that $T_t(\Delta) \to L_k$ for all k, so that $T_t(\Delta) \to \{\beta\}$.

We consider real-valued functionals of $T(t, \cdot)$. First we consider the net flow into the set $\{1, \cdots, s\}$, defined by

$$F_{st}(\mathbf{x}) = \sum_{j=1}^{s} T_j(t, \mathbf{x})$$

with derivative

$$F_s'(\mathbf{x}) = \hat{\alpha}(1 - x_s)^m - \sum_{j=1}^{s} jx_j, \tag{63}$$

which is continuous and strictly decreasing in $\mathbf{x}$. Moreover, for all $\mathbf{x}$ sufficiently large, $F_s'(\mathbf{x}) < 0$; and for all $\mathbf{x}$ sufficiently small, $F_s'(\mathbf{x}) > 0$. Consequently, $F_{st}(\mathbf{x}) \to 0$, $F_s'[T(t, \mathbf{x})] \to 0$ and $T_t(\Delta) \to L_1$ as $t \to \infty$, where

$$L_1 = \{\mathbf{x} \in \Delta: F_s'(\mathbf{x}) = 0\}. \tag{64}$$

Next consider the net flow into the states $\{1, \cdots, s-1\}$, defined by

$$F_{(s-1),t}(\mathbf{x}) = \sum_{j=1}^{s-1} T_j(t, \mathbf{x})$$

with derivative

$$\begin{aligned} F_{s-1}'(\mathbf{x}) &= \hat{\alpha}(1 - x_s)^{m-1}(1 - x_s - x_{s-1}) - \sum_{j=1}^{s-1} jx_j \\ &= \left[\hat{\alpha}(1 - x_s)^m - \sum_{j=1}^{s} jx_j\right] + [sx_s - x_{s-1}\hat{\alpha}(1 - x_s)^{m-1}] \\ &= F_s'(\mathbf{x}) + [sx_s - x_{s-1}\hat{\alpha}(1 - x_s)^{m-1}]. \end{aligned} \tag{65}$$

For $\mathbf{x} \in L_1$, $F_s'(\mathbf{x}) = 0$, and $F_{s-1}'(\mathbf{x}) = [sx_x - x_{s-1}\hat{\alpha}(1 - x_s)^{m-1}]$, which is continuous and strictly decreasing in $(x_{s-1}, -x_s)$. For all $\mathbf{x} \in L_1$ with x_{s-1} sufficiently large (small) and x_s sufficiently small (large), $F_{s-1}'(\mathbf{x}) < 0$ (>0). Hence, $F_{(s-1)t}(\mathbf{x}) \to 0$ and $F_{s-1}'[T(t, \mathbf{x})] \to 0$ for $\mathbf{x} \in L_1$, and $T_t(L_1) \to L_2$ as $t \to \infty$, where

$$L_2 = \{\mathbf{x} \in L_1: F_{s-1}'(\mathbf{x}) = 0\}. \tag{66}$$

Similarly, we consider the net flow $F_{(s-2)t}(\mathbf{x})$ into the states $\{1, \cdots, s-2\}$ with derivative

$$F'_{s-2}(\mathbf{x}) = F'_s(\mathbf{x}) + F'_{s-1}(\mathbf{x}) + (s-1)x_{s-1} - x_{s-2}\hat{\alpha}(1-x_s)^{m-1}, \quad (67)$$

which is continuous and strictly decreasing in $(x_{s-2}, x_{s-1}, -x_s)$. Moreover, for all $\mathbf{x} \in L_2$ with x_{s-2} sufficiently large (small) and x_{s-1} and x_s sufficiently small (large), $F'_{s-2}(\mathbf{x}) < 0$ (>0). Hence, $T_t(L_2) \to L_3$, where

$$L_3 = \{\mathbf{x} \in L_2 : F'_{s-2}(\mathbf{x}) = 0\}. \quad (68)$$

The proof is completed by induction. The s equations $F'_k(\mathbf{x}) = 0$, $1 \le k \le s$, uniquely determine the fixed point β of $T(t, \cdot)$ in Δ defined by (10) and (11). These are the partial balance equations for a single M/M/s/loss facility.[6] Hence, $L_s = \{\beta\}$ and $T_t(\Delta) \to \{\beta\}$ as $t \to \infty$. □

3.3 Proof of Theorem 3(a)

Proof: We now apply Theorems 8 and 9 to prove Theorem 3(a). Let $\mathbf{Z}_n = (Z_{n1}, \cdots, Z_{ns})$ have the unique stationary distribution of $\{\mathbf{X}_n(t), t \ge 0\}$ for each n. (Existence and uniqueness follow from Theorem 4.) Since the state space for $\{\mathbf{Z}_n\}$ is the compact simplex Δ in R^s, the sequence $\{\mathbf{Z}_n\}$ is uniformly tight and has a weakly convergent subsequence, say $\{\mathbf{Z}_{n_k}\}$; apply the argument in the proof of Theorem 8. Since $\mathbf{Z}_{n_k} \Rightarrow \mathbf{Z}$ in Δ as $n_k \to \infty$ for some $\mathbf{Z}$, the stationary versions of the stochastic processes $\mathbf{X}_{n_k}(t)$ converge weakly (in distribution) in $D([0, \infty), \Delta)$ and $n_k \to \infty$ to the continuous deterministic motion $\mathbf{X}(t)$ with $\mathbf{X}(0)$ distributed as $\mathbf{Z}$ (applying Theorem 8). However, since $\mathbf{X}_{n_k}(t)$ is stationary for each n_k, so is $\mathbf{X}(t)$. By Corollary 9.1, the only stationary distribution for $\mathbf{X}(t)$ is the limiting vector β. Hence, we must have $P(\mathbf{Z} = \beta) = 1$. Since every convergent subsequence of $\{\mathbf{Z}_n\}$ has the same limit $\mathbf{Z}$, we must have convergence of the entire sequence, that is, $\mathbf{Z}_n \Rightarrow \mathbf{Z}$ in Δ (see Theorem 2.3 of Ref. 61). Since $P(\mathbf{Z} = \beta) = 1$ for the deterministic vector β, we have convergence in probability (see page 25 of Ref. 61). □

Remark: Theorems 3, 8, and 9 together imply that the stationary versions of the stochastic processes $\mathbf{X}_n(t)$ also satisfy a functional law of large numbers in $D([0, \infty), \Delta)$.

3.4 Proof of Theorem 3(b)

The key to Theorem 3(b), of course, is Theorem 3(a) and the symmetry: Every subset of size m is equally likely to be the set of m required facilities for each arrival. In addition to Theorem 3(b) we establish a stronger form of asymptotic independence, for the stochastic processes instead of only the stationary distributions. Let $Y_{ni}(t)$ be the number of busy servers at facility i at time t. Let $Y_n(t) = [Y_{n1}(t), \cdots, Y_{nn}(t)]$ be the stationary version for each n.

Theorem 10: For any finite subset H and any t_0, the stationary stochastic processes $\{Y_{ni}(t), 0 \le t \le t_0\}$, $i \in H$, are asymptotically independent as $n \to \infty$.

Proof: By symmetry, the joint distribution of $\{Y_{ni}(t), i \in H\}$ is invariant under a permutation of the indices. By Theorem 3a, the proportion of facilities with j busy servers converges in probability to β_j as $n \to \infty$. Hence, by symmetry, $\lim_{n\to\infty} P[Y_{ni}(0) = j_i, 1 \le i \le H] = \prod_{i \in I} \beta_{j_i}$, so that the initial stationary values $Y_{ni}(0)$, $i \in I$, are asymptotically mutually independent. Next, let $A_{ni}(t)$ be the arrival process to facility i excluding losses due to blocking elsewhere. By Theorem 3a, $A_{ni}(t)$ converges to a Poisson process with rate $\hat{\alpha}(1 - \beta_s)^{m-1}$ as $n \to \infty$. Moreover, again by symmetry and Theorem 3a, the arrival processes $\{A_{ni}(t), 0 \le t \le t_0\}$, $i \in H$, are asymptotically mutually independent as $n \to \infty$. Since probability that the facilities in H share any customers at any time in the interval $[0, t_0]$ is asymptotically negligible as $n \to \infty$, the departure processes for $i \in H$ and thus also the processes $\{Y_{ni}(t), 0 \le t \le t_0\}$, $i \in H$, are asymptotically mutually independent. □

IV. EXISTENCE, UNIQUENESS, AND INSENSITIVITY

We now prove Theorem 4.

Proof: In the case of exponentially distributed service times, the vector-valued stochastic process, say $[N_1(t), \cdots, N_c(t)]$, representing the number of class j customers in service at time t for all j, $1 \le j \le c$, is an irreducible c-dimensional continuous-time Markov chain with a finite state space. Hence, there exists a unique stationary distribution. It is easy to see that the claimed distribution in Theorem 4 is the steady-state distribution by making the standard partial balance analysis.[6-8] The same steady-state distribution holds for general service-time distributions by the insensitivity results, which we discuss further below.

To prove the rest of Theorem 4, we need to establish that the steady-state distribution of $(N_1, \cdots, N_c)$ is actually well defined. For this purpose, we construct a continuous-time vector-valued Markov process $\{\mathbf{Z}(t), t \ge 0\}$, depicting the number of class j customers in service for each j and the remaining service time of each at time t. [$\mathbf{Z}(t)$ is the continuous-time Markov process associated with the GSMP in Ref. 46.] The steady-state distribution in Theorem 4 is understood to be the marginal distribution corresponding to $(N_1, \cdots, N_c)$ of the stationary distribution of $\mathbf{Z}(t)$. We shall show that $\mathbf{Z}(t)$ indeed has a stationary distribution (without establishing uniqueness) and that the marginal distribution corresponding to $(N_1, \cdots, N_c)$ is always as claimed in Theorem 4 and so is unique.

For our given general service-time distributions, we construct se-

quences of approximating service-time distributions from finite mixtures of finite convolutions of exponential distributions, as in Section 3.3 of Ref. 6 and in the proof of Theorem 2 in Ref. 46. We construct this so that the means are unchanged and there is weak convergence to the given distributions. For our special model, it is easy to see that each continuous-time Markov process $\mathbf{Z}(t)$ so created with these approximating service-time distributions has a unique invariant probability measure. Existence follows from the theory of continuous-time Markov chains with finite-state space. Uniqueness follows from the irreducibility that is evident from our special structure. Moreover, the partial balance property satisfied by the steady-state distribution in the exponential case implies that the unique stationary distribution of $\mathbf{Z}(t)$ in each approximating case has marginal distribution for $(N_1, \cdots, N_c)$ as specified in Theorem 4.[8,44,45] Finally, we treat the case of the original general service-time distributions by continuity, invoking Theorem 3 of Ref. 46. (Note that uniqueness with the approximating service-time distributions is crucial for that theorem.) This continuity theorem implies that the process $\mathbf{Z}(t)$ indeed has a stationary distribution and that the marginal distribution corresponding to $(N_1, \cdots, N_c)$ is as claimed for every stationary distribution of $\mathbf{Z}(t)$. □

Remark: An alternate proof of existence and uniqueness can be constructed using the fact that arrival epochs when the system is empty constitute regeneration points. The GSMP theory is also useful for describing steady state in more general models for which this is not the case; for example, if the service-time distributions are nonexponential and the arrival processes for the different classes are independent non-Poisson renewal processes. However, the insensitivity is typically lost with this extension.

V. CONVERGENCE OF THE SUCCESSIVE APPROXIMATION ALGORITHM

Example 3 in Section 1.5 showed that the successive approximation scheme (8) need not converge. In this section we show that if the offered loads are sufficiently small, then the operator T defined by the right side of (7) is a contraction operator, so that it has a unique fixed point to which successive iterates of T converge geometrically fast. However, the conditions for this property are quite strong, so that the theorem does not nearly cover all practical cases.

To state our results, let $\|\cdot\|$ be the supremum norm on R^n defined by $\|\mathbf{x}\| = \max\{|x_i| : 1 \le i \le n\}$ for $\mathbf{x} = (x_1, \cdots, x_n)$. Let $\tilde{\alpha}_i(\mathbf{b})$ be the reduced offered load as a function of $\mathbf{b} \equiv (b_1, \cdots, b_n)$ as defined in (6). Let $\gamma(\mathbf{b})$ be defined by

$$\gamma(\mathbf{b}) = \max_{1 \le i \le n} \left\{ \left(\frac{s_i}{\tilde{\alpha}_i(\mathbf{b})} - 1 + b_i \right) b_i n \hat{\alpha}_i \right\}. \tag{69}$$

Let $\mathbf{U} \equiv (U_i, \cdots, U_n)$ be an upper bound on any solution $\mathbf{b}^*$ of (7) such as $(B(s_1, \hat{\alpha}_1), \cdots, B(s_n, \hat{\alpha}_n)) = T^2(1)$ or $T^{2k}(1)$ for any $k \geq 1$.

Theorem 11: If $\gamma(\mathbf{U}) < 1$ *for* γ *in* (69) *and the upper bound* $\mathbf{U}$ *to any solution of* (7), *then*

(*i*) $\| T(\mathbf{b}^1) - T(\mathbf{b}^2) \| \leq \gamma(\mathbf{U}) \| \mathbf{b}^1 - \mathbf{b}^2 \|$ *for all* $\mathbf{b}^1$ *and* $\mathbf{b}^2$ *in* R^n *with* $0 \leq b_i^1, b_i^2 \leq U_i$ *for all* i, *so that*

(*ii*) T *has a unique fixed point* $\mathbf{b}^*$ *in* $[\mathbf{0}, \mathbf{U}] \equiv \{\mathbf{b}: 0 \leq b_i \leq U_i\}$, *and*

(*iii*) $\| T^k(\mathbf{b}^0) - \mathbf{b}^* \| \leq \gamma(\mathbf{U})^k \| \mathbf{b}^0 - \mathbf{b}^* \|$ *for all* k *when the initial vector* $T^0(\mathbf{b}) = \mathbf{b}^0$ *is in* $[\mathbf{0}, \mathbf{U}]$.

Proof: Parts (ii) and (iii) follow from (i) by the Banach-Picard fixed-point theorem for a contraction map on a complete metric space.[70] For (i) it suffices to have

$$\left| \frac{\partial T_i(\mathbf{b})}{\partial b_k} \right| \leq \frac{\gamma(\mathbf{U})}{n}$$

for all i and k (for example, see Theorem 2, page 111 of Ref. 70). By Theorem 15 of Jagerman,[27]

$$\frac{\partial B(s, \alpha)}{\partial \alpha} = \left[\frac{s}{\alpha} - 1 + B(s, \alpha) \right] B(s, \alpha).$$

Hence,

$$\left| \frac{\partial T_i(\mathbf{b})}{\partial b_k} \right| \leq \left[\frac{s_i}{\bar{\alpha}_i(\mathbf{b})} - 1 + b_i \right] b_i \hat{\alpha}_i \leq \frac{\gamma_i(\mathbf{U})}{n}$$

for $\mathbf{b} \in [\mathbf{0}, \mathbf{U}]$.

Remark 1: For the symmetric model, (69) simplifies to

$$\gamma(U) = \frac{nsU}{(1 - U)^{m-1}} - (1 - U)Un\hat{\alpha}, \tag{70}$$

so that a simple sufficient condition for the condition of theorem 11 is

$$\frac{nsU}{(1 - U)^{m-1}} < 1. \tag{71}$$

Remark 2: If $U_i = B(s_i, \hat{\alpha}_i)$ for all i or if $\mathbf{U} = T^{2k}(1)$ for some k using (8), then U is an increasing function of the offered loads $(\hat{\alpha}_1, \cdots, \hat{\alpha}_n)$ or $(\alpha_1, \cdots, \alpha_c)$ because T is an increasing function of $\mathbf{b}$ and $B(s, \alpha)$ is an increasing function of α. Hence, if the offered loads are sufficiently small, then the vector $\mathbf{U}$ will be sufficiently small, so that the condition of Theorem 11 will eventually hold.

VI. CONCLUSIONS

We have investigated a model to describe the blocking probabilities

when service is required from several multiserver facilities simultaneously. We have shown in Theorem 1 that some standard approximations produce upper bounds. In the process, we have established several other useful stochastic comparison results (Theorems 5 through 7 and Ref. 31). We also have proposed an improved reduced-load approximation and developed an efficient algorithm (Theorem 2) to treat both the Poisson arrival case (Section 1.5) and the non-Poisson arrival case (Section 1.9). In Theorem 8 we have established a functional law of large numbers that implies that the symmetric reduced-load approximation is asymptotically correct for symmetric models as the number of facilities increases with the offered load per facility and the number of facilities per class held fixed (Theorems 3 and 10). We have displayed the exact formula in Theorem 4 and justified the insensitivity with respect to the service-time distributions (Sections 1.8 and IV).

Among the important directions for future research are (i) testing the approximations further, especially for non-Poisson arrival processes; (ii) establishing better conditions for the reduced-load equations (7) to have a unique solution (Conjectures 1 and 2); (iii) establishing better conditions for the successive approximation scheme (8) to converge; (iv) establishing lower bounds on the exact blocking probabilities paralleling the upper bounds in Theorem 1; (v) determining if the reduced-load approximation is an upper bound on the exact blocking probability for symmetric models (Conjecture 3); (vi) determining if the exact blocking probabilities for symmetric models are increasing in n when the offered load per facility is fixed (Conjecture 4); (vii) establishing (if possible) the diffusion limit in Section III (Conjectures 5 and 6); (viii) seriously analyzing smaller models in which the basic facility-independence approximation in (5) underlying all the approximations here is not appropriate.[9] In particular, in the spirit of Kaufman[7] and Mitra and Weinberger,[21] it would be nice to develop an efficient algorithm for the exact blocking probabilities in Theorem 4 and Corollary 4.1.

It would also be of interest to consider other related models, for example, models in which more than one server per facility may be required, and related delay systems. There are two kinds of waiting to be considered for delay systems: waiting for each customer class outside the system, and waiting for service at each facility within the system. The second form of waiting may still require simultaneous service or some other form.[14]

VII. ACKNOWLEDGMENTS

This work was initially motivated by discussions with D. D. Sheng about the PANDA software package.[2,3] I am grateful to her, D. P.

Heyman (Bell Communications Research), and D. R. Smith for initial stimulating discussions. The product bound in Theorem 1 was also conjectured by them. The summation bound in Corollary 1.2 was proved in the special case of two facilities by different methods by Sheng and Smith (see the appendix of Ref. 3). Symmetric models were apparently first investigated by Mitra and Weinberger,[21] but they were brought to my attention by Heyman and Smith. Heyman and Smith also developed the reduced-load approximation (9) for symmetric models and conjectured Theorem 3. I am grateful to W. A. Massey for showing me his unpublished work[38] and for commenting on Ref. 31, which presents the general stochastic comparison theory behind Theorems 1, 6, and 7. Finally, I am grateful to W. J. Hery and J. T. Wittbold (AT&T Communications) for discussions about their applications and the performance of the approximations.[24,25] They each programmed the reduced-load approximation (5) through (8) and investigated numerical examples. Tables IV and V come from Wittbold.

VIII. EPILOGUE

This section has been added in proof to report important new work. Kelly[71] has proved that the reduced-load system of eq. (7) has a unique solution, thus confirming Conjectures 1 and 2. Kelly also has proved that the reduced-load approximation is asymptotically correct in heavy traffic, that is, in a network with fixed topology in which $\alpha_j \to \infty$ and $s_i \to \infty$, as in Ref. 57. In fact, Kelly's heavy-traffic limit theorem is a multifacility generalization of the local limit theorem in the Appendix of Ref. 57.

Ziedens and Kelly[72] also have proved limit theorems similar to Theorem 3 for symmetric networks in which the number of nodes increases. For the special tree networks in Fig. 1, Mitra[73] has determined an efficient algorithm for the exact solution based on asymptotic expansions, in the spirit of Ref. 21. Other related work appears in Refs. 74 through 76.

REFERENCES

1. D. Bear, *Principles of Telecommunication-Traffic Engineering*, London: The Institution of Electrical Engineers, 1976.
2. D. D. Sheng, "Performance Analysis Methodology for Packet Network Design," IEEE Global Telecommun. Conf., *GLOBECOM '83* (December 1983), pp. 456–60.
3. C. L. Monma and D. D. Sheng, unpublished work.
4. E. Cinlar, *Introduction to Stochastic Processes*, Englewood Cliffs, New Jersey: Prentice-Hall, 1975.
5. R. W. Wolff, "Poisson Arrivals See Time Averages," Oper. Res., *30,* No. 2 (March–April 1982), pp. 223–31.
6. F. P. Kelly, *Reversibility in Stochastic Networks*, New York: Wiley, 1979.
7. J. S. Kaufman, "Blocking in a Shared Resource Environment," IEEE Trans. Commun., *COM-29,* No. 10 (October 1981), pp. 1474–81.
8. D. Y. Burman, J. P. Lehoczky, and Y. Lim, "Insensitivity of Blocking Probabilities

in a Circuit-Switching Network," J. Appl. Probab., *21,* No. 4 (December 1984), pp. 850–59.
9. J. M. Holtzman, "Analysis of Dependence Effects in Telephone Trunking Networks," B.S.T.J., *50,* No. 8 (October 1971), pp. 2647–62.
10. V. E. Beneš, "Models and Problems of Dynamic Memory Allocation," *Applied Probability—Computer Science: The Interface,* Vol. I, ed. R. L. Disney and T. J. Ott, Boston: Birkhauser, 1982, pp. 89–135.
11. E. G. Coffman, Jr., T. T. Kadota, and L. A. Shepp, "A Stochastic Model of Fragmentation in Dynamic Storage Allocation," SIAM J. Comput., *14,* No. 2 (May 1985), pp. 416–25.
12. G. F. Newell, "The M/M/∞ Service System With Ranked Servers in Heavy Traffic," Lecture Notes in Economics and Math. Systems, *231,* New York: Springer-Verlag, 1984.
13. L. A. Gimpelson, "Analysis of Mixtures of Wide and Narrow Band Traffic," IEEE Trans. Commun. Technol., *13* (1965), pp. 258–66.
14. E. Wolman, "The Camp-On Problem for Multiple-Address Traffic," B.S.T.J., *51,* No. 6 (July–August 1972), pp. 1363–422.
15. K. J. Omahen, "Capacity Bounds for Multiresource Queues," J. Assoc. Comput. Mach., *24,* No. 4 (October 1977), pp. 646–63.
16. E. Arthurs and J. S. Kaufman, "Sizing a Message Store Subject to Blocking Criteria," *Performance of Computer Systems,* ed. M. Arato, A. Butrimenko, and E. Gelenbe, Amsterdam: North-Holland, 1979, pp. 547–64.
17. L. Green, "A Queueing System in Which Customers Require a Random Number of Servers," Oper. Res., *28,* No. 6 (November–December 1980), pp. 1335–46.
18. P. H. Brill and L. Green, "Queues in Which Customers Receive Simultaneous Service From a Random Number of Servers: A System Point Approach," Manage. Sci., *30,* No. 1 (January 1984), pp. 51–68.
19. L. Green, "A Multiple Dispatch Queueing Model of Police Patrol Operations," Manage. Sci., *30,* No. 6 (June 1984), pp. 653–64.
20. A. Federgruen and L. Green, "An M/G/c Queue in Which the Number of Servers Required is Random," J. Appl. Probab., *21,* No. 3 (September 1984), pp. 583–601.
21. D. Mitra and P. J. Weinberger, "Probabilistic Models of Database Locking: Solutions, Computational Algorithms, and Asymptotics," J. Assoc. Comput. Mach., *31,* No. 4 (October 1984), pp. 855–78.
22. D. Mitra, unpublished work.
23. D. P. Heyman, "Asymptotic Marginal Independence in Large Networks of Loss Systems," Bell Communications Research, Holmdel, 1985. Presented at the ORSA/TIMS Applied Probability Conf., Williamsburg, Va., January 1985.
24. W. J. Hery, private communication.
25. J. T. Wittbold, AT&T Communications, private communication.
26. J. M. Akinpelu, "The Overload Performance of Engineered Networks With Nonhierarchial and Hierarachial Routing," AT&T Bell Lab. Tech. J., *63,* No. 7 (September 1984), pp. 1261–82.
27. D. L. Jagerman, "Some Properties of the Erlang Loss Functions," B.S.T.J., *53,* No. 3 (March 1974), pp. 525–51.
28. D. L. Jagerman, "Methods in Traffic Calculations," AT&T Bell Lab. Tech. J., *63,* No. 7 (September 1984), pp. 1283–310.
29. W. Feller, *An Introduction to Probability Theory and Its Applications,* Vol. I, Third Edition, New York: Wiley, 1968.
30. D. R. Smith and W. Whitt, "Resource Sharing for Efficiency in Traffic Systems," B.S.T.J., *60,* No. 1 (January 1981), pp. 39–55.
31. W. Whitt, unpublished work.
32. D. J. Daley, "Stochastically Monotone Markov Chains," Zeitschrift Wahrscheinlichkeitstheorie Verw. Geb., *10* (1968), pp. 305–17.
33. J. Keilson and A. Kester, "Monotone Matrices and Monotone Markov Processes," Stoch. Proc. Appl., *5,* No. 3 (July 1977), pp. 231–41.
34. A. Kester, *Preservation of Cone Characterizing Properties of Markov Chains,* Ph.D. Thesis, University of Rochester, 1977.
35. D. Stoyan, *Comparison Methods for Queues and Other Stochastic Models,* ed. D. J. Daley, New York: Wiley, 1983.
36. W. A. Massey, "An Operator Analytic Approach to the Jackson Network," J. Appl. Probab., *2* (June 1984), pp. 379–93.
37. W. A. Massey, "Open Networks of Queues: Their Algebraic Structure and Estimating Their Transient Behavior," Adv. Appl. Probab., *16,* No. 1 (March 1984), pp. 176–201.
38. W. A. Massey, unpublished work.

39. N. Dunford and J. T. Schwartz, *Linear Operators, Part I: General Theory*, New York: Interscience, 1958.
40. W. Whitt, "Open and Closed Models for Networks of Queues," AT&T Bell Lab. Tech. J., *63*, No. 9 (November 1984), pp. 1911–79.
41. E. Fuchs and P. E. Jackson, "Estimates of Distributions of Random Variables for Certain Communications Models," Commun. ACM, *13*, No. 12 (December 1970), pp. 752–7.
42. P. F. Pawlita, "Traffic Measurements in Data Networks, Recent Measurement Results, and Some Implications," IEEE Trans. Commun., *COM-29*, No. 4 (April 1981), pp. 525–35.
43. W. T. Marshall and S. P. Morgan, unpublished work.
44. R. Schassberger, "Insensitivity of Steady-State Distributions of Generalized Semi-Markov Processes With Speeds," Adv. Appl. Probab., *10*, No. 4 (December 1978), pp. 836–51.
45. D. Y. Burman, "Insensitivity in Queueing Systems," Adv. Appl. Probab., *13*, No. 4 (December 1981), pp. 846–59.
46. W. Whitt, "Continuity of Generalized Semi-Markov Processes," Math. Oper. Res., *5*, No. 4 (November 1980), pp. 494–501.
47. E. Brockmeyer, H. L. Halstrom, and A. Jensen (eds.), *The Life and Works of A. K. Erlang*, Copenhagen: Danish Academy of Sciences, 1948.
48. F. Baskett et al., "Open, Closed and Mixed Networks of Queues With Different Classes of Customers," J. Assoc. Comput. Mach., *22*, No. 2 (April 1975), pp. 248–60.
49. F. P. Kelly, "Networks of Queues," Adv. Appl. Prob., *8*, No. 2 (June 1976), pp. 416–23.
50. R. Schassberger, "The Insensitivity of Stationary Probabilities in Networks of Queues," Adv. Appl. Probab., *10*, No. 4 (December 1978), pp. 906–12.
51. K. Matthes, "Zur Theorie der Bedienungsprozesse," Trans. Third Prague Conf. Inf. Theory, Prague, 1962.
52. A. D. Barbour, "Networks of Queues and the Method of Stages," Adv. Appl. Probab., *8*, No. 3 (September 1976), pp. 584–91.
53. S. S. Lam, "Queueing Networks With Population Size Constraints," IBM J. Res. Develop., *21*, No. 4 (July 1977), pp. 370–8.
54. P. Franken et al., *Queues and Point Processes*, Berlin: Akademie-Verlag, 1981.
55. A. E. Eckberg, "Generalized Peakedness of Teletraffic Processes," Proc. Tenth Int. Teletraffic Congress, Montreal, June 1983, p. 4.4 b.3.
56. A. A. Fredericks, "Approximating Parcel Blocking via State Dependent Birth Rates," Proc. Tenth Int. Teletraffic Congress, Montreal, June 1983, p. 5.3.2.
57. W. Whitt, "Heavy-Traffic Approximations for Service Systems With Blocking," AT&T Bell Lab. Tech. J., *63*, No. 5 (May–June 1984), pp. 689–708.
58. R. E. Barlow and F. Proschan, *Statistical Theory of Reliability and Life Testing*, New York: Holt, Rinehart and Winston, 1975.
59. S. Karlin and Y. Rinott, "Classes of Orderings of Measures and Related Correlation Inequalities. I. Multivariate Totally Positive Distributions," J. Multivar. Anal., *10*, No. 4 (December 1980), pp. 467–98.
60. T. Kamae, U. Krengel, and G. L. O'Brien, "Stochastic Inequalities on Partially Ordered Space," Ann. Probab., *5*, No. 6 (December 1977), pp. 899–912.
61. P. Billingsley, *Convergence of Probability Measures*, New York: Wiley, 1968.
62. W. Whitt, "On the Heavy-Traffic Limit Theorem for GI/G/∞ Queues," Adv. Appl. Probab., *14*, No. 1 (March 1982), pp. 171–90.
63. D. W. Stroock and S. R. S. Varadhan, *Multidimensional Diffusion Processes*, New York: Springer-Verlag, 1979.
64. T. Lindvall, "Weak Convergence of Probability Measures and Random Functions in the Function Space $D[0, \infty)$," J. Appl. Probab., *1* (March 1973), pp. 109–21.
65. W. Whitt, "Some Useful Functions for Functional Limit Theorems," Math. Oper. Res., *5*, No. 1 (February 1980), pp. 67–85.
66. L. Arnold, *Stochastic Differential Equations: Theory and Applications*, New York: Wiley, 1974.
67. W. Whitt, "Weak Convergence of Probability Measures on the Function Space $C[0, \infty)$," Ann. Math. Statist., *41*, No. 3 (June 1970), pp. 939–44.
68. K. R. Parthasarathy, *Probability Measures on Metric Spaces*, New York: Academic Press, 1967.
69. W. Whitt, "Comparing Counting Processes and Queues," Adv. Appl. Probab., *13*, No. 1 (March 1981), pp. 207–20.
70. E. Isaacson and H. B. Keller, *Analysis of Numerical Methods*, New York: Wiley, 1966.

71. F. P. Kelly, "Blocking Probabilities in Large Circuit-Switched Networks," Statistical Laboratory, University of Cambridge, England, 1985.
72. I. B. Ziedens and F. P. Kelly, "Loss Probabilities in Circuit-Switched Star Networks," Statistical Laboratory, University of Cambridge, England, 1985.
73. D. Mitra, unpublished work.
74. P. M. Lin et al., "Analysis of Circuit-Switched Networks Employing Originating Office Control With Spill Forward," IEEE Trans. Commun., *COM-26,* No. 6 (June 1978), pp. 754–65.
75. A. Girard and Y. Ouimet, "End-to-End Blocking for Circuit-Switched Networks: Polynomial Algorithms for Some Special Cases," IEEE Trans. Commun., *COM-31,* No. 12 (December 1983), pp. 1269–73.
76. G. Iazolla, P. J. Courtois, and A. Hordijk, *Mathematical Computer Performance and Reliability,* Amsterdam: North-Holland, 1984.

AUTHOR

Ward Whitt, A.B. (Mathematics), 1964, Dartmouth College; Ph.D. (Operations Research), 1968, Cornell University; Stanford University, 1968–1969; Yale University, 1969–1977; AT&T Bell Laboratories, 1977—. At Yale University, from 1973–1977, Mr. Whitt was Associate Professor in the departments of Administrative Sciences and Statistics. At AT&T Bell Laboratories he is in the Operations Research Department of the Systems Analysis Center, where the primary mission is to investigate and improve the product realization process.

AT&T Technical Journal
Vol. 64, No. 8, October 1985
Printed in U.S.A.

Performance Comparison of InGaAsP Lasers Emitting at 1.3 and 1.55 μm for Lightwave System Applications

By N. K. DUTTA,* R. B. WILSON,† D. P. WILT,* P. BESOMI,‡ R. L. BROWN,* R. J. NELSON,† and R. W. DIXON*

(Manuscript received April 11, 1985)

Experimental results relative to the performances of real index-guided InGaAsP lasers emitting near 1.3 and 1.55 μm are described and compared. The laser structures discussed are the etched mesa buried heterostructure, channeled substrate buried heterostructure, and the double channel planar buried heterostructure. The effect of Auger recombination and intervalence band absorption on the threshold current and external differential quantum efficiency is discussed. The effect of the larger Auger coefficient at 1.55 μm is compensated by a lower carrier density at threshold at 1.55 μm so that the total nonradiative current loss for lasers emitting at 1.55 μm is not significantly larger than that for lasers emitting at 1.3 μm. A small linear shunt leakage current (~10 mA) can increase the T_0 to ~100K. We report threshold currents as low as 11 and 15 mA (at 30°C) and continuous-wave operating temperatures as high as 130 and 110°C for lasers emitting at 1.3 and 1.55 μm, respectively.

I. INTRODUCTION

Lightwave transmission systems are being installed throughout the world at a rapidly escalating pace. These new systems offer higher bit rate and longer repeater spacing than conventional systems and thus reduce the transmission cost per bit.[1]

* AT&T Bell Laboratories. † Lytel, Inc., Somerville, New Jersey. ‡ General Optronics, Edison, New Jersey.

Most of the lightwave systems installed so far (often called first-generation systems) use multimode fibers and an operating wavelength of about 0.85 μm.[2] Second-generation systems using single-mode fibers and an operating wavelength near 1.3 μm offer longer repeater spacings because the loss of silica fibers is lower at 1.3 μm than at 0.85 μm.[2,3] Zero chromatic dispersion near 1.3 μm for silica fibers allows the use of multimode laser sources in single-mode fibers for long-distance high bit-rate transmission without significant dispersion penalty.[2,3] Although the full potential of lightwave systems operating at 1.3 μm has not yet been realized because it is a relatively new technology, it is quite conceivable that a third-generation system operating at 1.55 μm may be extensively installed in the near future because the silica fiber loss is minimum at 1.55 μm.[4] Laboratory experiments have already demonstrated the high system performance that can be achieved at this wavelength.[5,6]

Development and subsequent installation of new lightwave systems have been driven by the development of new high-quality components, namely fibers, sources (lasers), detectors (avalanche photodiode or pin photodiode), or integrated front ends, and transmitter and receiver electronics. For example, the development of low-cost silica fibers with zero dispersion at 1.55 μm will certainly influence the third-generation systems. High-speed integrated circuits for receiver and transmitter packages are currently being developed for high bit-rate (>1 Gb/s) systems. This paper compares the performance of InGaAsP lasers emitting at 1.3 and 1.55 μm. The former is currently in use in several second-generation systems and the latter can be a single-frequency source for third-generation systems operating at that wavelength using conventional single-mode silica fibers.

The performance requirements for lasers used in lightwave transmitters usually include linear (kink-free) light-current characteristics up to a certain power output (typically ~5-mW/facet, ~0-dBm power input to the fiber), capability of high bit-rate modulation, and long operating life. Low-threshold and high-differential quantum efficiency are desirable (necessary for some systems) in order to reduce the bias current and modulation needed from the drive circuitry. Since most lightwave transmitters need to operate over a range of temperature, a weak temperature dependence of the threshold of the laser is desirable—otherwise, thermoelectric controllers are required inside the transmitter package in order to stabilize the laser temperature. The conventional InGaAsP double heterostructure laser emitting at 1.55 μm may be more sensitive to temperature and have lower differential quantum efficiency than a laser emitting at 1.3 μm. The former is due to larger nonradiative Auger recombination rate[7,8] and the latter may be due to larger intervalence band absorption[8,9] at longer wavelengths.

The effect of these processes is discussed in detail in Sections II and III. We find that the effect of the larger Auger coefficient at 1.55 μm is compensated by smaller carrier density at threshold so that the nonradiative Auger current at 1.55 μm is not significantly larger than that for 1.3 μm.

The fabrication of InGaAsP lasers emitting at 1.55 μm by Liquid Phase Epitaxy (LPE) usually requires the growth of an additional InGaAsP layer (antimeltback layer) in order to prevent the meltback of the active layer in subsequent growth of InP layer.[10]

The effect of the thickness of this antimeltback layer on threshold current and efficiency of 1.55-μm lasers is discussed in Section IV. Although this layer need not be present when the double heterostructure is grown by Vapor Phase Epitaxy (VPE), fabrication of Distributed Feedback (DFB)-type single-frequency lasers usually requires an intermediate band gap layer (1.1 to 1.3 μm) between the active layer (1.55 μm) and InP layers. Antimeltback layers are not needed for 1.3-μm InGaAsP double heterostructures grown by LPE, although DFB lasers emitting at 1.3 μm need an intermediate gap layer (~1.1 μm) between the active layer and the InP cladding layers for optimization.

Real index-guided lasers are needed as sources for high-performance fiber communication systems because these lasers are less susceptible to light-current nonlinearities and intensity self-pulsations than gain-guided lasers.[11] Many strongly index-guided laser structures utilize reverse biased junctions for current confinement. Leakage currents, that is, current flowing around the active region, may be responsible for high-threshold and light-current sublinearity in nonoptimized structures.[12] Since the leakage currents in many cases varies as $\exp(\Delta E_g/kT)$, where ΔE_g is the difference in band gap of the blocking layers (InP) and the active region (1.3- or 1.55-μm InGaAsP), we expect the leakage currents in 1.55-μm InGaAsP lasers to be smaller than those in 1.3-μm InGaAsP lasers. This is discussed in Section V.

Experimental results from several types of real index-guided lasers emitting at 1.3 and 1.55 μm are compared in Section VI. These results show that the 1.55-μm lasers have somewhat higher threshold current (~30 percent at 30°C) and lower light output at 100 mA (~40 percent at 30°C). The former, in our opinion, is due principally to smaller mode confinement factor (because of the presence of antimeltback layer), and the latter is due to the combined effect of lower photon energy (~20 percent) and larger intervalence band absorption and free carrier absorption at 1.55 μm. The measured threshold current as a function of temperature for lasers emitting at both wavelengths can be represented by the expression $I_{th} \sim I_0 \exp(T/T_0)$, where T_0 is a parameter determining the temperature sensitivity. Similar T_0 values are observed for lasers emitting at both wavelengths except in cases

where leakage current is believed to affect T_0. The limitations in high bit-rate long-haul fiber communication systems introduced by the source linewidth are discussed in Section VII.

II. AUGER EFFECT

Since the initial work by Beattie and Landsberg,[13] it has been established that the band-to-band Auger processes are often a major nonradiative carrier loss mechanism in small band gap semiconductors. We expect the nonradiative Auger rate for 1.55-μm band gap material to be larger than that for 1.3-μm material.

The Auger rate (R_a) in undoped semiconductor varies approximately as

$$R_a = \gamma n^3, \tag{1}$$

where γ is the Auger coefficient and n is the injected carrier density.[13] The radiative recombination rate varies approximately as

$$R_r = Bn^2, \tag{2}$$

where B is the radiative recombination coefficient.[14,15] The current density of a broad-area laser at lasing threshold (in the absence of other recombination mechanisms) is given by

$$\begin{aligned} J &= eR_r d + eR_a d \\ &= J_r + J_a, \end{aligned} \tag{3}$$

where e is the electron charge and d is the active layer thickness, J_r, J_a are the radiative and the Auger component of the current, respectively. The Auger coefficient γ in eq. (1) is dominated by phonon-assisted Auger processes for larger band gap semiconductors and by band-to-band processes for small band gap semiconductors. Detailed discussion of band-to-band, phonon-assisted, and trap Auger processes in direct gap semiconductors are given in Ref. 7. Figure 1 shows the calculated γ for InGaAsP alloy lattice matched to InP for $n = 10^{18}$ cm^{-3}. Because of the uncertainty in the calculation of the absolute magnitude of the Auger coefficient, we have plotted γ in Fig. 1 normalized to its value (γ_0) for 1.3-μm InGaAsP. The calculated γ_0 $\sim 1 \times 10^{-28}$ cm^6 sec^{-1}. The measured values by several authors are shown in Table I. Figure 1 shows that the Auger coefficient for 1.55-μm InGaAsP is about a factor of 4 larger than that for 1.3-μm InGaAsP. Agrawal and Dutta[20] have found that the Auger coefficient for 1.55-μm InGaAsP is about a factor of 3 larger than that for 1.3-μm InGaAsP from an analysis of threshold current of stripe geometry InGaAsP lasers emitting at 1.3 and 1.55 μm. Thus the Auger coefficient for 1.55-μm InGaAsP is about 3 to 4 times larger than that for 1.3-μm InGaAsP.

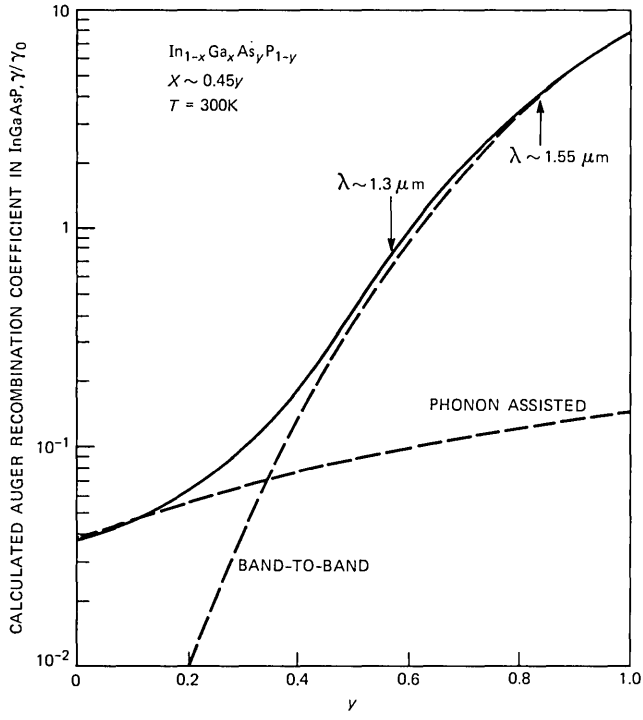


Fig. 1—The relative Auger coefficient γ/γ_0 is plotted as a function of band gap of InGaAsP.

Table I—Measured Auger coefficients

(cm^6sec^{-1})	Reference	Comment
5×10^{-29} (CCHS)	Mozer et al. (1982)[16]	$\lambda = 1.3\ \mu m$
$2.3 \pm 1 \times 10^{-29}$ (Total)	Sermage et al. (1983)[17]	$\lambda = 1.3\ \mu m$, optically pumped
1×10^{-29} (Total)	Henry et al. (1981)	$\lambda = 1.3\ \mu m$, doping dependence of τ
$<3 \times 10^{-29}$ (Total)	Su et al. (1982)	$\lambda = 1.3\ \mu m$
3×10^{-29} (Total)	Thompson (1983)[18]	Fit to data of Su et al.
$3\text{–}8 \times 10^{-29}$ (Total)	Uji (1983)[19]	$\lambda = 1.3\ \mu m$ LEDs
$2.8\text{–}1.5 \times 10^{-29}$	Wintner and Ippen (1984)	$\lambda = 1.3\ \mu m$ optically pumped
7.5×10^{-29}	Wintner and Ippen (1984)	$\lambda = 1.55\ \mu m$ optically pumped

The radiative recombination rate B can be calculated by using the Gaussian Halperin-Lax band tails and Stern's matrix elements.[14,15] Figure 2 shows the calculated B/B_0 for InGaAsP for $n = 10^{18}$ cm^{-3}. B_0 is the value of B for 1.3-μm InGaAsP. The calculated $B_0 = 1 \times 10^{-10}$ cm^3 sec^{-1}. The measured values lie in the range $0.9\text{–}1.5 \times 10^{-10}$ cm^3 sec^{-1}.[21,22] Equations (1) and (2) are strictly valid only for nondegenerate electron and hole gas. However, at high injected carrier densities ($\sim 2 \times 10^{18}$ cm^{-3}) at laser threshold the electron and holes are degen-

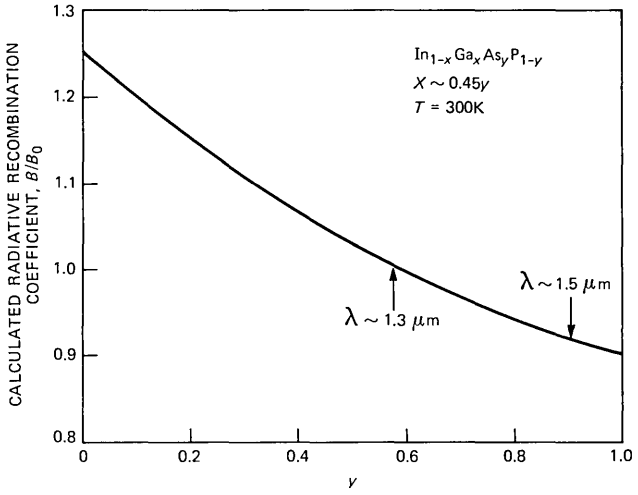


Fig. 2—The relative radiative recombination coefficient B/B_0 is plotted as a function of band gap of InGaAsP.

erate. Degeneracy effects can be introduced by writing γ, B as functions of n. Both γ and B decrease slowly with increasing carrier density.[7,8,14,15]

The laser threshold currents are determined principally by the magnitudes of γ, B, and the threshold carrier density. The latter depends to some extent on the laser structure, for example, through the mode confinement factor and optical absorption. However, it is possible to determine the injected carrier density at transparency (n_0) from the band structure parameters alone.[23] We consider undoped materials. Using the Joyce-Dixon approximation[24] for the quasi-Fermi levels and assuming parabolic bands, we show the calculated n_0 for InGaAsP in Fig. 3. The threshold carrier density is usually 30 to 40 percent higher than n_0. Note that n_0 for 1.55-μm InGaAsP (in Fig. 3) is smaller than that for 1.3-μm InGaAsP. This is due to smaller conduction band effective mass at long wavelengths. Since the radiative recombination current varies as Bn^2, Figs. 2 and 3 suggest that the threshold current of 1.55-μm InGaAsP lasers should be lower than that for 1.3-μm InGaAsP lasers in the absence of Auger recombination. This is shown in Fig. 4c.

Figure 4 shows the calculated radiative and Auger components of the current, plotted as a function of temperature, for 1.3- and 1.55-μm InGaAsP-InP broad-area double heterostructure lasers. The radiative component is calculated using a constant (temperature-independent) absorption loss of 30 cm^{-1} in the active layer as in Ref. 7. The Auger component of the total current is calculated using eq. (1) for Auger

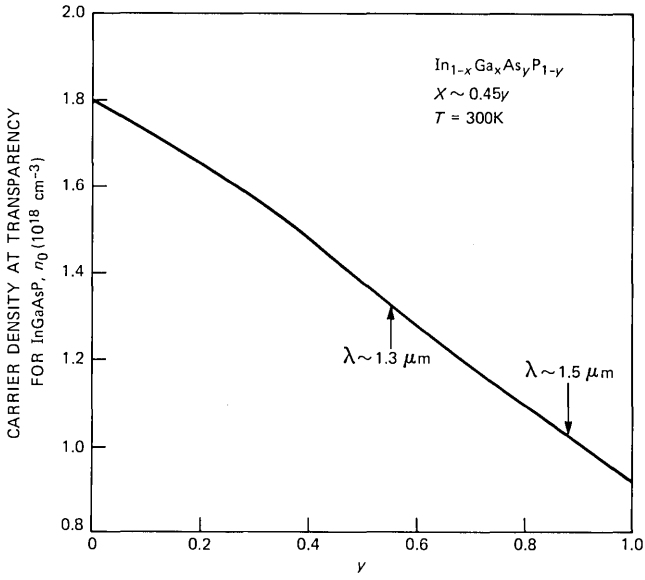


Fig. 3—The injected carrier density at transparency for undoped InGaAsP.

recombination rate, the temperature dependence of threshold carrier density (n_{th}) from Figs. 17 and 23 of Ref. 7, the calculated temperature dependence of γ, and $\gamma_0 = 3 \times 10^{-29}$ cm^6 sec^{-1} (which falls within the range of the measured values but is smaller than the calculated value[7] of 1×10^{-28} cm^6 sec^{-1}).

Auger rate calculations using the Kane band model[25] result in Auger coefficients higher than observed experimentally.[7,8] Recently, Haug[26] has calculated the Auger coefficients in 1.3-μm InGaAsP using the band model of Chelikowsky and Cohen[27] for InP but with an energy gap assumed to be that of γ = 1.3-μm InGaAsP. He finds that the phonon-assisted Auger processes (CCCHP[7]) are dominant with a value of $\sim 2.5 \times 10^{-29}$ cm^6 sec^{-1}. Using a calculated temperature dependence of the phonon-assisted Auger process, this would result in a T_0 value of 100 to 110K for 1.3-μm InGaAsP-InP lasers, which is higher than the observed value in broad-area lasers (~60 to 70K). There is sufficient uncertainty in the band structure parameters and the carrier concentration at threshold that it is unreasonable to expect better agreement between calculation and experiment. Figure 5 shows the predicted T_0 values (between 300 and 350K) plotted as a function of Auger coefficient (at 300K) for two different values of carrier density at threshold. The quantity k is the ratio of the Auger coefficients at 350 and 300K. The smaller value is the calculated result for the phonon-assisted Auger processes, and the higher value is that for the band-to-band processes. Although the temperature dependence of γ

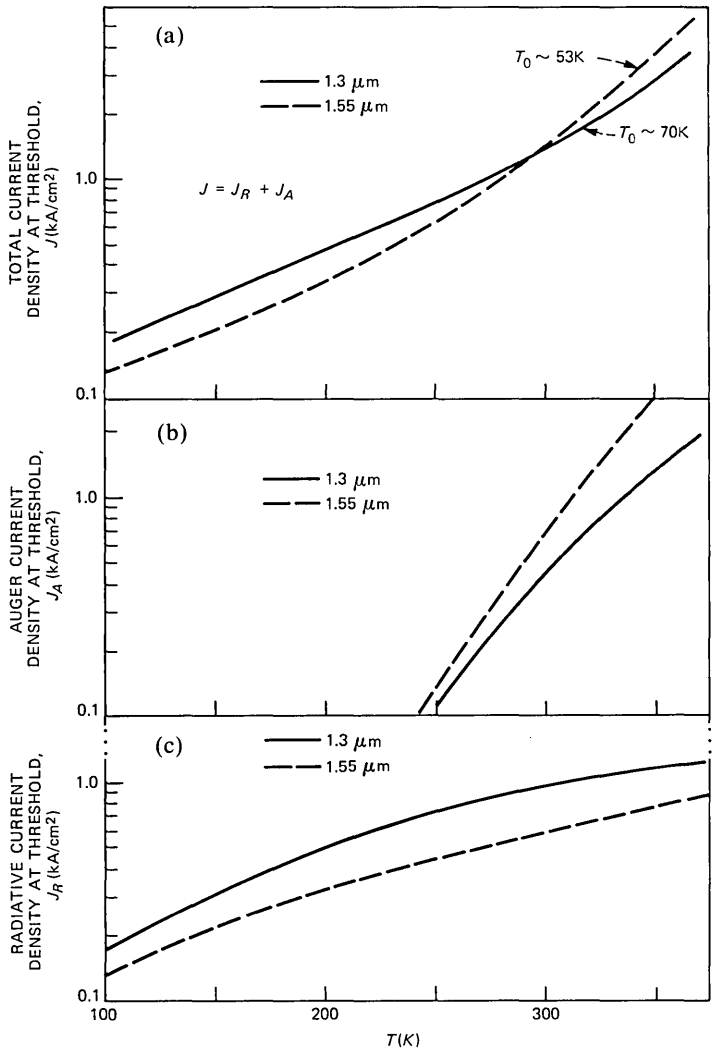


Fig. 4—The calculated radiative and Auger component of the current for lasers emitting at 1.3 and 1.55 μm. (a) Radiative. (b) Auger component. (c) Total.

has not been experimentally determined, the calculated temperature dependence of n_{th} agrees well with that of the measured threshold current using short pulses.[28] The total threshold current which is the sum of the radiative and Auger component is shown in Fig. 4a. The smaller J_R for 1.55-μm lasers is approximately compensated by the larger J_A. The above calculation is for an InGaAsP active layer with InP cladding layers both for 1.3- and 1.55-μm lasers. The presence of an antimeltback layer reduces the confinement factor, which increases

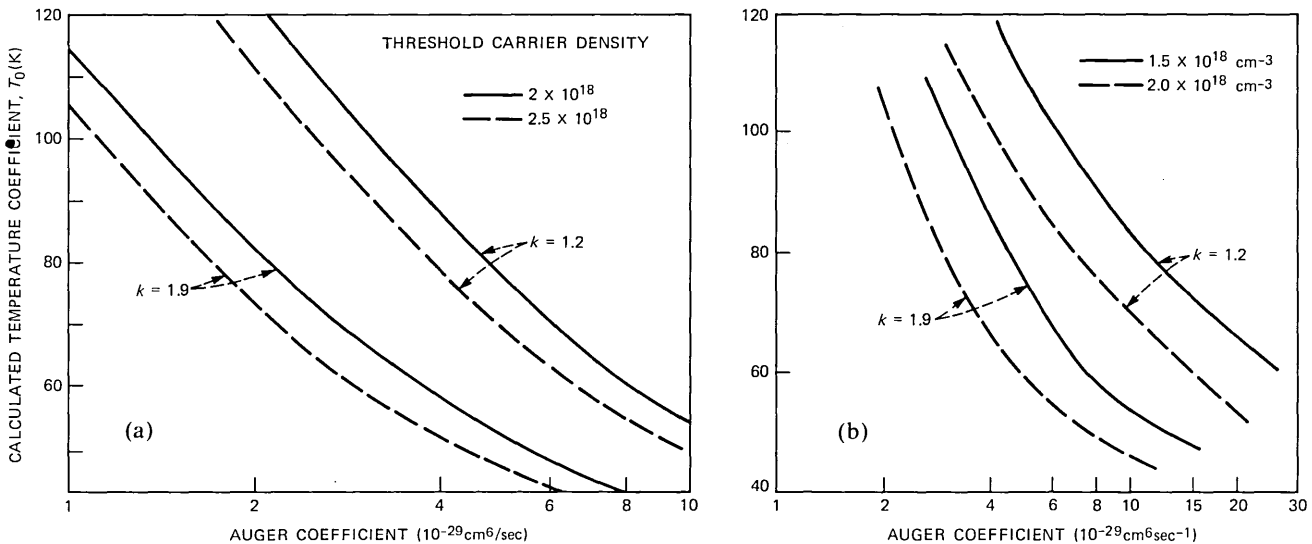


Fig. 5—The calculated T_0 values as a function of Auger coefficient. (a) λ = 1.3-μm InGaAsP. (b) λ = 1.55-μm InGaAsP.

the threshold current of 1.55-μm lasers over that shown in Fig. 4. This is discussed in detail in Section IV. Furthermore, heterobarrier leakage due to drift and diffusion may be responsible for increased threshold current of 1.3-μm InGaAsP DH lasers if the p-cladding layer doping is too low.[29,30] These lasers also exhibit a higher temperature dependence of threshold (lower T_0) than is commonly observed. Such heterobarrier leakage is smaller in 1.55-μm InGaAsP lasers because of larger barrier height. The high-energy electrons generated in the Auger process may also escape (thermionic emission) over the heterobarrier, in which case the carrier leakage is expected to be independent of the barrier height and p-cladding doping. Shah et al.[31] have reported observation of hot carriers, and Yamakoshi et al.[32] have reported observation of carrier leakage in 1.3-μm InGaAsP-InP lasers. However, Henry et al.[33] did not observe hot carriers in their experiments.

III. INTERVALENCE BAND ABSORPTION

Intervalence band absorption is an optical loss mechanism in the active region of InGaAsP lasers that can reduce the external differential quantum efficiency.[8,9] The external differential quantum efficiency (η_d) is given approximately by

$$\eta_d = \frac{\alpha_m}{\alpha_m + \alpha}, \tag{4}$$

where

$$\alpha = \Gamma\alpha_a + (1 - \Gamma)\alpha_c,$$

where α_m is the mirror loss, Γ is the confinement factor, and α_a and α_c are the absorption and cladding layer losses, respectively. For a 250-μm long laser the "equivalent distributed" loss $\alpha_m \simeq 40$ cm^{-1} using R = 0.35. Figure 6 shows η_d plotted as a function of α_a for several values of Γ. We assume $\alpha_c = 30$ cm^{-1}. The light output L at a current ΔI above threshold is given by $L = \eta_d E_g \Delta I$, where E_g is the band gap of the active layer. The light output from 1.55-μm lasers for a given ΔI and η_d is lower than that for 1.3-μm lasers because the former has lower photon energy.

Sugimura[8] has calculated the intervalence band absorption in III–V semiconductors using Kane band model. The absorption is larger for small band gap semiconductors and increases with increasing temperature. The calculated absorption (α) normalized to its value (α_0) for 1.3-μm InGaAsP at 300K is shown in Fig. 7. The calculated α_0 is approximately 30 cm^{-1}. Henry et al.[34] have extrapolated the intervalence band absorption in 1.55-μm InGaAsP from measurements in InGaAs. Adams et al.[9] have proposed that the high temperature

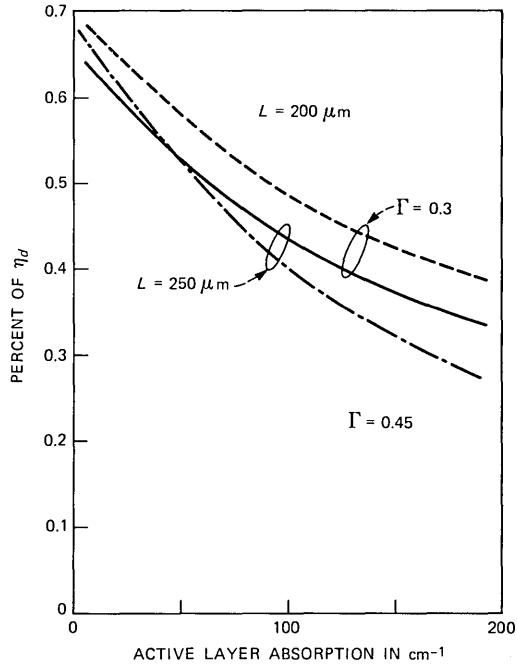


Fig. 6—The calculated external differential quantum efficiency as a function of optical loss in the active.

sensitivity of threshold (low T_0) current of 1.6-μm InGaAsP lasers is due to intervalence band absorption. Although this is not the dominant mechanism responsible for low T_0 of 1.55-μm InGaAsP lasers, it may be why the observed η_d of InGaAsP lasers emitting at 1.55 μm is lower than that for lasers emitting at 1.3 μm.

IV. ANTIMELTBACK THICKNESS

Both 1.3- and 1.55-μm InGaAsP double heterostructures have been grown by LPE and VPE techniques. However, as mentioned previously, for LPE growth it is necessary to grow a short-wavelength (~1.1 to 1.3-μm) InGaAsP layer over the 1.55-μm active layer to prevent meltback of the active layer during the subsequent growth of InP layer. A thick antimeltback layer reduces the confinement factor of the guided mode and hence increases the laser threshold. However, a smaller confinement factor reduces the effect of intervalence band absorption and hence increases the differential quantum efficiency. We now calculate the effect of antimeltback layer thickness on device threshold.

The threshold gain g_{th} is given by

$$\Gamma g_{\text{th}} = \Gamma\alpha_a + (1 - \Gamma)\alpha_c, \tag{5}$$

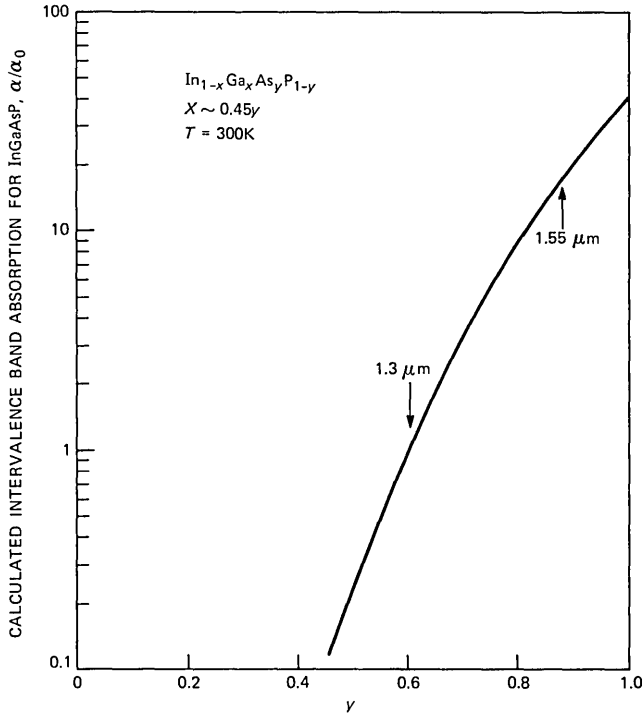


Fig. 7—The calculated intervalence band absorption from Ref. 8 normalized to its value for 1.3 μm.

where Γ, α_a, and α_c are the confinement factor and the absorption in the active and cladding layers, respectively. We assume $\alpha_c = 30$ cm^{-1} and $\alpha_a = 40$ cm^{-1} and 150 cm^{-1} for the curves in Fig. 8. The confinement factor is calculated using the analysis of Butler[35] for a four-layer guide. The refractive index of InP, 1.3-μm InGaAsP, and the 1.55-μm active layer are 3.21, 3.4, and 3.55, respectively. For the calculated g_{th} in eq. (5), the threshold carrier density (n_{th}) is obtained from previous calculations. Then the threshold current density at 300K is obtained using eq. (3) with $B = 0.9 \times 10^{-10}$ cm^{-3} sec^{-1} and $\gamma = 9 \times 10^{-29}$ cm^6 sec^{-1}. Figure 8 shows the results of the calculation for both J_{th} and η_d as a function of antimeltback layer thickness. Thus, both the threshold current and the external differential quantum efficiency increase as the antimeltback layer thickness is increased. The former is due to larger threshold gain caused by reduced confinement factor as the antimeltback layer thickness is increased, and the latter is due to lower loss experienced by the lasing mode in the active layer (in the above mode) due to smaller Γ. The observed smaller η_d of the 1.55-μm lasers may also be due to larger free carrier absorption both in the cladding and active layer at longer wavelengths.

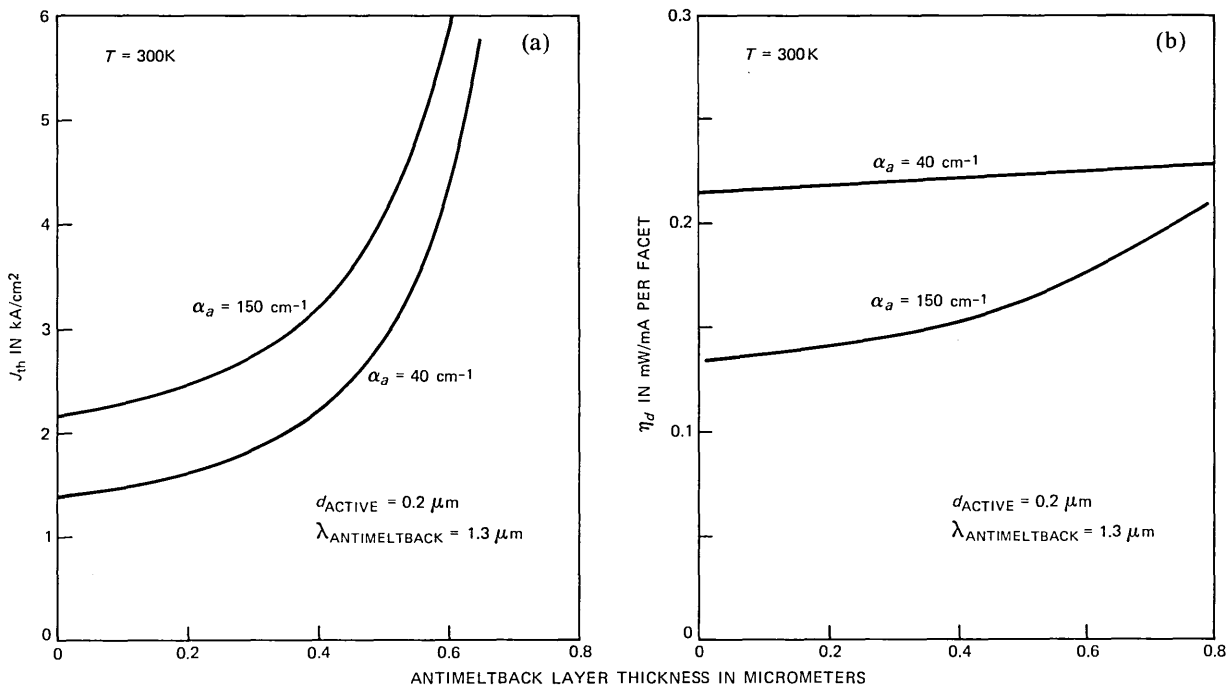


Fig. 8—The (a) threshold current and (b) external differential quantum efficiency as a function of the thickness of the antimeltback layer.

V. LEAKAGE CURRENT

Real index-guided lasers are needed for high bit-rate fiber transmission systems because of their superior performance over gain-guided lasers.[11] Many strongly index-guided lasers utilize reverse biased junctions for current confinement to the active region.[12] It is generally believed that leakage currents, that is, currents flowing around the active region, are responsible for high threshold current, light-current sublinearity, and poor high-temperature performance of nonoptimized laser structures. We have previously analyzed the leakage currents in several 1.3-μm InGaAsP-InP laser structures using electrical equivalent circuit models.[12] The leakage current through leakage paths with pin junctions varies approximately as $\exp(\Delta E_g/kT)$, where ΔE_g is the band gap difference between InP and the active layer. $\Delta E_g \sim 0.39$ eV and ~0.55 eV for InGaAsP lasers emitting at 1.3 and 1.55 μm, respectively. Thus the magnitude of leakage current is expected to be approximately $\sim\exp(0.16\ \text{eV}/kT) \sim e^6$ times smaller for InGaAsP lasers emitting at 1.55 μm than for lasers emitting at 1.3 μm. In many laser structures the leakage current flows through forward-biased pin InP homojunctions.[12] A fraction of this current in 1.3-μm InGaAsP lasers can be detected as radiative emission at ~0.95 μm, which is the band gap of InP.[36] No emission at ~0.95 μm is observed from 1.5-μm InGaAsP lasers, which suggests that leakage currents through InP homojunctions are significantly smaller in these lasers. These considerations do not apply to linear resistive shunt paths.

VI. EXPERIMENTAL RESULTS

In this section, we compare the experimental results from several types of real index-guided InGaAsP lasers emitting at 1.3 and 1.55 μm. We first discuss the strongly index-guided lasers. Over the last few years, we have fabricated the Channeled Substrate Buried Heterostructure (CSBH),[37] the Etched Mesa Buried Heterostructure (EMBH)[38,39] and the Double-Channel Planar Buried Heterostructure (DCPBH)[40] lasers (see Fig. 9).[41] The CSBH laser has a nonplanar active region and can be fabricated using one LPE growth. The EMBH and DCPBH lasers have planar active regions, which make them compatible with the fabrication of DFB and DBR-type single-frequency lasers.[42] These structures need two epitaxial growth steps for fabrication. Schematic cross sections of these laser structures are shown in Fig. 9.

6.1 Double channel planar buried heterostructure lasers

The schematic cross section of this device structure is shown in Fig. 9c. The fabrication of DCPBH lasers involves two epitaxial growth

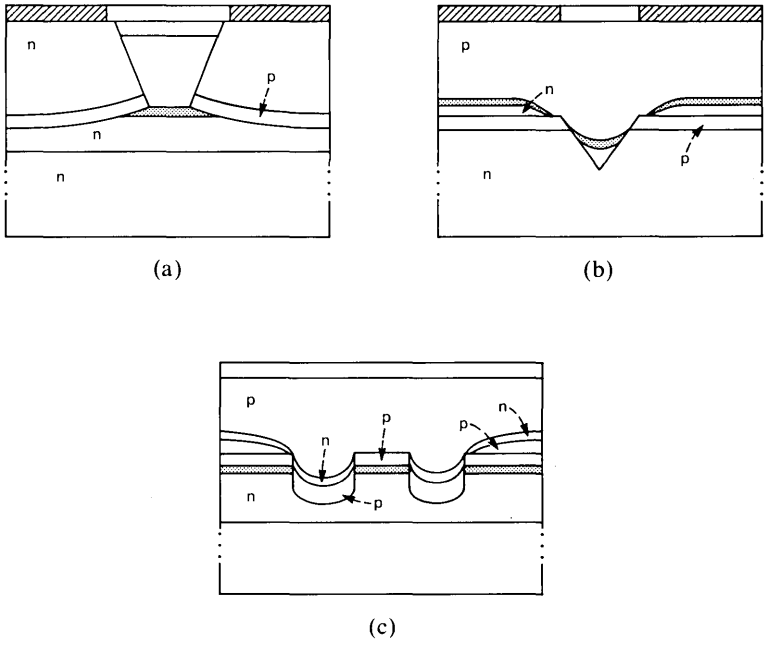


Fig. 9—The schematic cross sections of the (a) EMBH, (b) CSBH, and (c) DCPBH lasers are shown.

steps.[40] Briefly, a planar double heterostructure is grown by LPE over a n-InP substrate followed by the double channel etching and then regrowth. The CW light-current (L-I) characteristics of a DCPBH laser emitting at 1.3 μm is shown in Fig. 10. Note that this particular laser operates CW up to 130°C. These lasers typically have pulsed threshold currents in the range 15 to 25 mA at 30°C, and external differential quantum efficiencies in the range 0.2 to 0.25 mW/mA/facet. The variation of threshold current (I_{th}) with temperature (T) is given by the commonly used expression $I_{th}(T) \sim I_0 \exp(T/T_0)$, with T_0 values in the range 80 to 100K. The higher T_0 values of many DCPBH lasers when compared with broad area lasers ($T_0 \sim$ 60 to 70K) could be due to a small (~10 mA) temperature-independent leakage current. This is supported by our observation that lasers with low threshold (<15 mA at 30°C) have smaller T_0 values than lasers with somewhat higher threshold (~25 mA at 30°C).

Figure 11 shows the L-I characteristics of a DCPBH laser emitting at 1.55 μm. The doping levels and layer thicknesses of the lasers emitting at 1.55 and 1.3 μm are similar except that the 1.55-μm double heterostructure has an antimeltback layer (~0.15 μm thick) of 1.3-μm InGaAsP.

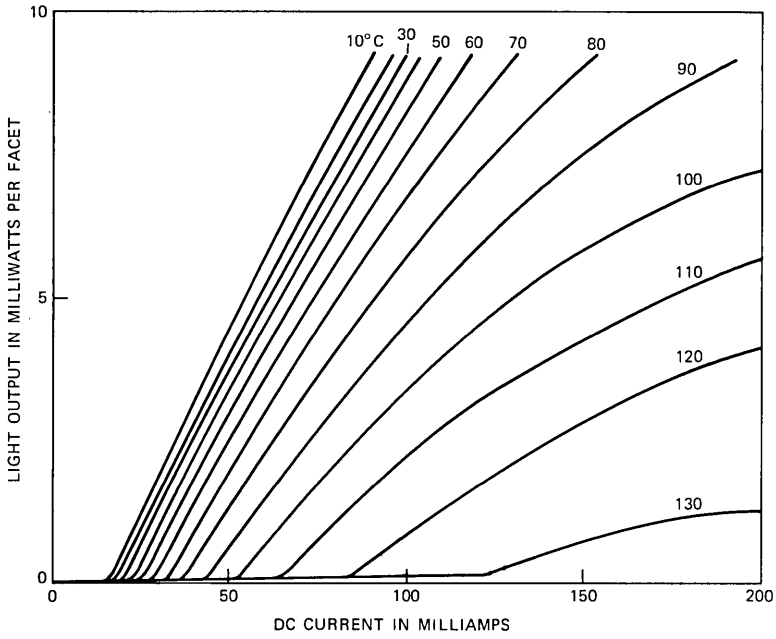


Fig. 10—Light-current characteristics of a DCPBH laser emitting at 1.3 μm.

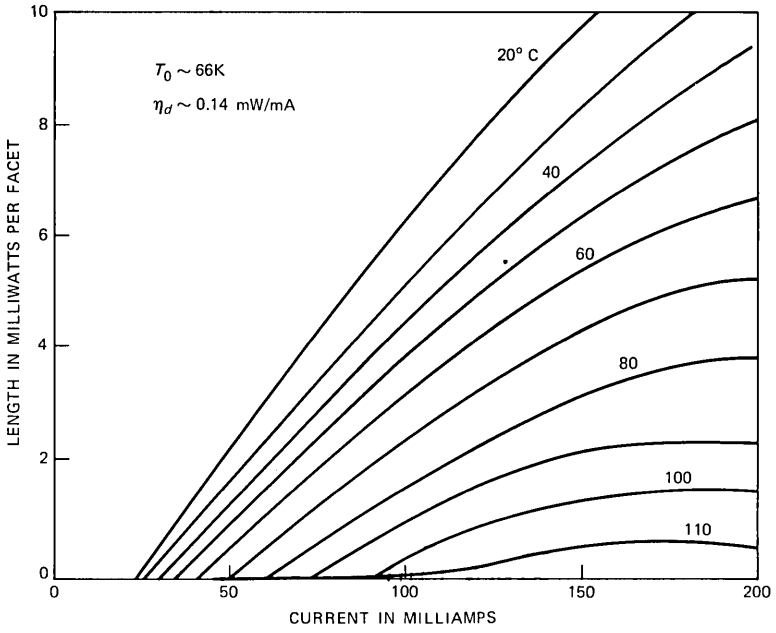


Fig. 11—Light-current characteristics of a DCPBH laser emitting at 1.55 μm.

Table II—Performance characteristics of typical DCPBH lasers

	$\lambda = 1.3\ \mu m$	$\lambda = 1.5\ \mu m$
I_{th} (30°C)	15–25 mA	20–35 mA
L(100 mA)	10–16 mW	6–9 mW
Efficiency/facet	0.2–0.25 mW/mA	0.12–0.15 mW/mA
T_0	90–100K	55–65K
T_{max} (CW)	130°C	110°C
Fabrication requirements		
Active area (kinks)	$<0.3\ \mu m^2$	$<0.2\ \mu m^2$
Antimeltback layer thickness (I_{th})	—	0.1–0.15 μm

The performance characteristics of our typical DCPBH lasers emitting at 1.3 and 1.5 μm are compared in Table II. The parameters shown are threshold current, external differential quantum efficiency, light output at 100 mA, T_0, and the maximum CW operating temperature (T_{max}). The performance of these lasers at both wavelengths are acceptable for many lightwave system applications. Nevertheless, we are interested in determining to what extent 1.3-μm lasers are intrinsically superior to 1.55-μm devices and to what extent the present differences represent relatively easy-to-overcome technological differences. The lowest threshold current observed for 1.3- and 1.55-μm lasers are 12 and 15 mA at 30°C. The maximum CW operating temperature of these lasers are 130 and 110°C, respectively.

The threshold current of the 1.5-μm lasers are slightly higher (~30 percent) than that for the 1.3-μm lasers. This is due to the smaller confinement factor of the waveguide mode of the lasers emitting at 1.5 μm caused by the antimeltback layer. This additional layer is not required for double heterostructures grown by vapor phase epitaxy, and hence in that case the threshold current may be lower.

The external differential quantum efficiency per facet of the lasers emitting at 1.55 μm is lower than that for lasers emitting at 1.3 μm. Using $\eta_d = 0.25$ mW/mA for 1.3-μm lasers and 0.15 mW/mA for 1.5-μm lasers, the optical absorption of the guided mode using eq. (4) is 37 and 67 cm^{-1} for the lasers emitting at 1.3 and 1.5 μm, respectively. In deriving the above, we have used a mirror loss of 40 cm^{-1}, which corresponds to a mode reflectivity of 0.35 for our 2.50-μm long lasers. The optical loss of the mode is related to the cladding and active layer loss by the following expression:

$$\alpha = \Gamma\alpha_a + (1 - \Gamma)\alpha_c. \tag{6}$$

Using $\alpha_c = 30\ cm^{-1}$ for both wavelengths and a calculated $\Gamma = 0.47$ and 0.38 for the 1.3- and 1.5-μm laser structure, we get $\alpha_a = 44$ and 127 cm^{-1} for the active layer losses of the 1.3- and 1.55-μm laser,

respectively, at 30°C. These estimates of the absorption loss are an upper limit because eq. (4) neglects the effect of leakage currents, which can reduce the efficiency. The external differential quantum efficiency can be increased by decreasing the length of the laser. For 1.55-μm lasers, a smaller confinement factor can also increase the efficiency, as discussed in Section IV.

The light output at 100 mA is also shown in Table II because it can be a measure of the combined effect of threshold current and external differential quantum efficiency and also provide an estimate of the drive current needed when the laser is used in lightwave transmitters. Note that this quantity is smaller for lasers emitting at 1.5 μm, because of the combined effect of lower efficiency and lower photon energy of these devices. Shorter lasers should have higher light output at 100 mA than that shown in Table II. However, the reliability of short lasers may be a problem because of higher threshold current (and carrier) density.

The observed temperature dependence of threshold current of lasers emitting at 1.3 μm is lower than that for lasers emitting at 1.5 μm. This is represented by higher T_0 value of the 1.3-μm lasers. We believe that a temperature-independent leakage path (leakage current ~10 mA at 30°C) can be responsible for high T_0 (~90 to 100K) of 1.3-μm lasers because low threshold (I_{th} ~ 15 mA at 30°C) 1.3-μm lasers have lower T_0 (~75K). High T_0 values caused by leakage current have been observed previously in EMBH lasers emitting at 1.3 μm.[43] The smaller leakage current in the DCPBH structure enables 1.3-μm DCPBH lasers to operate at temperatures as high as 130°C.

Spatial hole burning causes transverse mode transitions with increasing injection in strongly index-guided lasers.[44] These mode transitions appear as "kinks" in the *L-I* characteristics and are undesirable for lightwave applications. Spatial hole burning can be reduced and the *L-I* kinks eliminated by reducing the active area of the laser to less than ~0.3 μm^2 for lasers emitting at 1.3 μm[44] and ~0.2 μm^2 for lasers emitting at 1.55 μm. This fabrication requirement for kink-free operation of 1.55-μm lasers is more difficult to achieve.

6.2 Etched mesa buried heterostructure laser

The light-output-current characteristics of a EMBH laser emitting at 1.3 μm is shown in Fig. 12. These lasers typically have threshold current in the range 15 to 30 mA at 30°C and external differential quantum efficiency in the range 0.2 to 0.25 mW/mA. The lowest threshold current observed for 1.3-μm EMBH lasers is 11 mA at 30°C and the maximum CW operating temperature is 100°C. The variation of threshold current (I_{th}) with temperature (T) is given by T_0 values in the range 60 to 75K. These lasers with optimized layer thicknesses

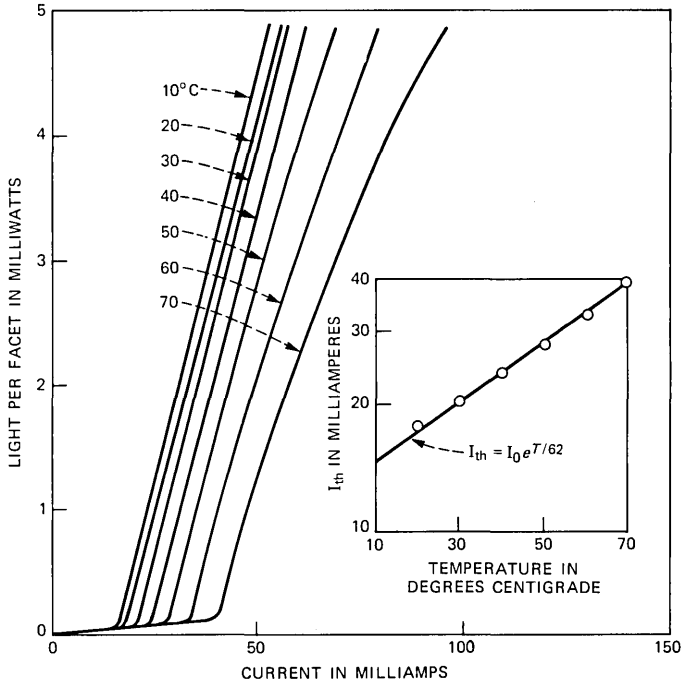


Fig. 12—Light-current characteristics of a EMBH laser emitting at 1.3 μm.

and doping levels do not exhibit a sublinearity in the light-current characteristics for power less than ~15 mW/facet near room temperature. At high injection, thyristor-like leakage path causes the *L-I* characteristics to roll over.[12,45] The fabrication requirements on active layer dimensions for low-threshold and kink-free operation of these lasers are shown in Table II.

6.3 Channeled substrate buried heterostructure lasers

We now discuss the experimental results for the CSBH laser (which has a nonplanar active layer). The schematic cross section of this device structure is shown in Fig. 9b. The light-current (*L-I*) characteristics of a CSBH laser emitting at 1.3 μm is shown in Fig. 13. The CSBH lasers are fabricated by LPE growth of an n-InP layer, 1.3-μm InGaAsP active layer, p-InP layer, and InGaAs contact layer over a base structure that has V grooves etched in it. The base structure has a p-InP current blocking layer, which may be LPE grown[37] or VPE grown[46] over an n-InP substrate. Cd-diffusion can also be used to form the blocking layer.[47] Alternatively, Fe implantation[48] or Fe-doped high-resistivity InP layers[49] can be used to limit the current flow to the active region in the V groove. All of the above schemes for base

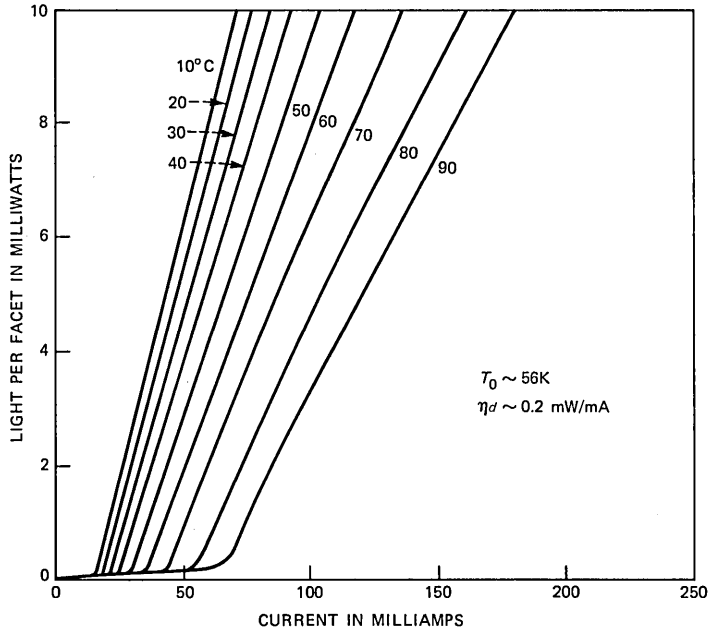


Fig. 13—Light-current characteristics of a CSBH laser emitting at 1.3 μm.

structures of CSBH lasers yield devices with comparable threshold current and efficiency. The threshold current of CSBH lasers emitting at 1.3 μm is typically in the range of 15 to 30 mA, and external differential quantum efficiency is 0.17 to 0.20 mW/mA/facet. Kink-free operation has been achieved to >24 mW/facet in lasers with active area less than 0.3 μm^2[50].

CSBH lasers emitting at 1.55 μm have been fabricated using Cd-diffused base structure and Fe-doped InP—grown by Metal Organic Chemical Vapor Deposition (MOCVD)—base structure. Figure 14 shows the *L-I* characteristics of a CSBH laser emitting at 1.55 μm with a Cd-diffused base structure. The CSBH lasers emitting at 1.55 μm have lower quantum efficiency than lasers emitting at 1.3 μm. The threshold current, quantum efficiency, and T_0 values of CSBH lasers emitting at 1.55 μm fabricated using Cd-diffused and MOCVD base structures are similar.

Performances of typical CSBH lasers emitting at 1.3 and 1.55 μm are compared in Table III. The parameters shown are threshold current, efficiency, light output at 100 mA, T_0, and maximum CW operating temperature. The data are representative of results from several wafers of each type. The observed lowest threshold currents for CSBH lasers emitting at 1.3 and 1.55 μm are 12 and 18 mA at

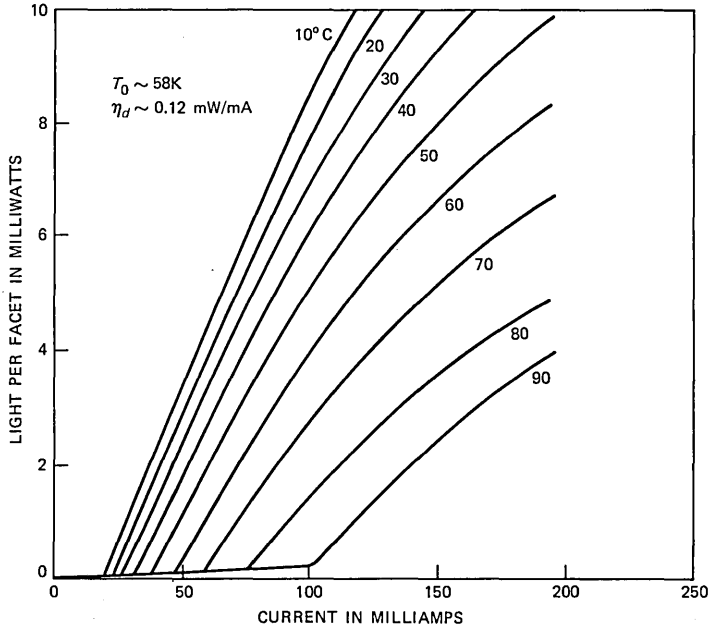


Fig. 14—Light-current characteristics of a CSBH laser emitting at 1.55 μm.

Table III—Performance characteristic of typical CSBH lasers

	$\lambda = 1.3$ μm	$\lambda = 1.5$ μm
I_{th} (30°C)	15–25 mA	25–35 mA
Efficiency/facet	0.17–0.2 mW/mA	0.11–0.14 mW/mA
L(100 mA)	8–13 mW	6–8 mW
T_0	50–55K	50K
T_{max}	90°C	90°C

30°C. The maximum CW operating temperatures for the 1.3- and 1.55-μm lasers are 120 and 90°C, respectively.

VII. DISCUSSION

The information-carrying capacity of a digital link is given by the product of the bit rate (B) and the distance (L) between the transmitter and the receiver. It is a common practice to characterize a transmission system by its bit-rate-distance product, although this may not be the judge of the overall performance.[51]

The loss limited transmission distance (L) is determined by the minimum number of photons per bit needed by a receiver to detect it. It is given by

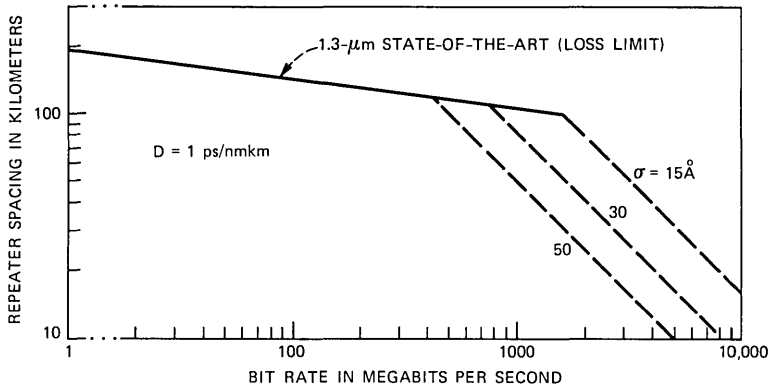


Fig. 15—Bit rate versus repeater spacing for lightwave systems operating near 1.3 μm.

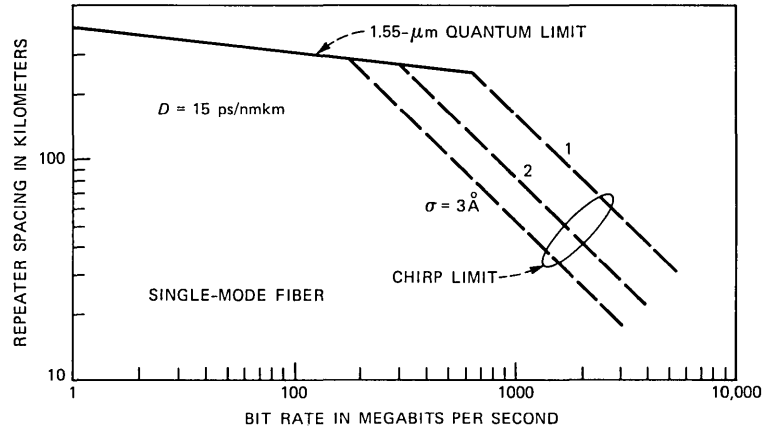


Fig. 16—Bit rate versus repeater spacing for lightwave systems operating near 1.55 μm.

$$P_R = P_T \exp(-\alpha L)$$

or

$$L = \frac{10}{\alpha} \log_{10} \frac{P_T}{P_R}, \tag{7}$$

where α is the fiber loss in dB/km and P_T, P_R is the transmitted and received power. For loss limit P_R varies linearly with the bit rate B. The solid lines in Figs. 15 and 16 show the loss limit for 1.3- and 1.55-μm systems using $P_T = 0$ dBm (1 mW).

At high bit-rate fiber dispersion becomes an important limitation because of pulse spreading. The effect of pulse spreading on the performance of lightwave systems have been calculated by several

authors.[52,53] The dispersion limit is usually given by[51] (within a factor of 2)

$$BL < 1/(AD\sigma), \tag{8}$$

where D is the fiber dispersion and σ is the source linewidth. Commercial silica fibers have zero dispersion near 1.3 μm and $D \sim 15$ ps/nmkm at 1.55 μm. Using $\sigma \sim 3$ nm for 1.55-μm multimode sources, eq. (8) suggests that in spite of the high loss limit the bit-rate-distance product is limited to 5.5 Gb-km for 1.55-μm transmission systems unless single-frequency sources are used.

Several schemes have been used to obtain single-frequency emission at 1.55 μm. Cleaved-coupled-cavity,[54] external cavity,[55] and DFB-type[42] single-frequency lasers have been fabricated using the DCPBH laser structure shown in Fig. 9c. Although all of these schemes produce essentially single-frequency sources (linewidth <20 MHz) under CW excitation, the linewidth under pulsed current modulation is significantly larger (1 to 2Å).[56-58] This phenomenon is usually called frequency chirping and it arises from a modulation of the refractive index by the injected current that modulates the effective cavity length. The chirp limit using eq. (8) for three different chirp widths is shown by the dashed curves in Fig. 16. External modulators must be used to eliminate the frequency chirping.

Dispersion limitations can also be important for 1.3-μm systems at high bit rates. In most optical fibers, the dispersion D is greater than 1 ps/nmkm at wavelength separation $\Delta\lambda > 10$ nm from the zero dispersion point. Using $D = 1$ ps/nmkm, we get the dashed lines in Fig. 15 as the dispersion limit for multimode source linewidths of 15, 30, and 50Å, respectively. The longitudinal mode spacing of a 2.50-μm long 1.3-μm InGaAsP laser is ~9Å. This suggests that the 30Å line in Fig. 9 may be a practical limit for 1.3-μm systems using multimode sources. Furthermore, it is well known that the emission spectrum of a 1.3-μm InGaAsP laser under microwave modulation (>2.5 Gb) is considerably broader (because of the appearance of many longitudinal modes) than for low bit-rate (<1 Gb) modulation.[59] This broadening is due to band filling, which becomes significant for modulation frequencies larger than inverse carrier recombination times. Thus, for high bit-rate-distance operation near 1.3 μm it is necessary to use single-frequency sources unless the emission spectrum under modulation is within ±10 nm of the zero dispersion wavelength of the fiber.

VIII. CONCLUSION AND SUMMARY

We have compared the performance of real index-guided InGaAsP lasers emitting at 1.3 and 1.55 μm. The 1.3-μm lasers have somewhat lower threshold current (~20 percent) than 1.55-μm lasers. This is

principally due to the smaller confinement factor of the 1.55-μm lasers due to the presence of the antimeltback layer in our LPE grown lasers. The VPE growth technique does not require the presence of antimeltbelt layer; thus essentially similar threshold current can be realized at both 1.3 and 1.55 μm. The measured median threshold current at 30°C from several wafers of our CSBH and DCPBH lasers are 15 to 25 and 20 to 35 mA for lasers emitting at 1.3 and 1.55 μm, respectively. The lowest threshold current observed at 30°C are 11 and 15 mA for lasers emitting at 1.3 and 1.55 μm, respectively.

The temperature dependence of the threshold current is given by the commonly used expression $I_{th}(\tau) \sim I_0 \exp(T/T_0)$ with T_0 ~60–75K for 1.3-μm lasers and 55- to 65K for 1.55-μm lasers. A small temperature-independent leakage current (~10 mA) is believed to be responsible for high T_0 values (~100K) of some 1.3-μm DCPBH lasers. The nonradiative Auger recombination process that increases with decreasing bandgap is believed to be responsible for the low T_0 values of the long-wavelength (both 1.3 and 1.55 μm) InGaAsP lasers. The effect of the larger Auger coefficient at 1.55 μm is compensated by lower carrier density at threshold at 1.55 μm so that the total nonradiative current loss for lasers emitting at 1.55 μm is not significantly larger than that for lasers emitting at 1.3 μm. This results in similar T_0 values for 1.3- and 1.55-μm lasers.

The measured efficiency of the 1.55-μm lasers is smaller (~20 percent) than that for 1.3-μm lasers. This may be due to the combined effect of larger intervalence band absorption and free carrier absorption at longer wavelengths. The smaller efficiency combined with lower photon energy (~20 percent) at 1.55 μm than at 1.33 μm makes the light output at a given operating current (~100 mA) of the 1.55-μm laser lower by ~35 percent than that for 1.3-μm lasers. This shows that the transmitter circuitry will need higher drive current when 1.55-μm lasers are used. The higher operating current may have some reliability implications. The external differential quantum efficiency for 1.55-μm lasers can be increased by fabricating short lasers. However, the reliability of short lasers may be a problem because of higher threshold current density in these devices.

Capacitance associated with leakage junctions is believed to influence the high frequency modulation capability of index-guided lasers that use reverse biased junctions for current confinement. We have fabricated 1.3- and 1.55-μm lasers that can be modulated at high bit rates (>2 Gb/s). The laser linewidth under modulation is found to be significantly broader than that under CW operation. The phenomenon is called frequency chirping and arises from the modulation of the carrier density that modulates the refractive index of the guided wave. The measured chirp width of the 1.55-μm laser is larger than that for

the 1.3-μm laser, because of the larger variation of refractive index with carrier density at longer wavelengths. The thickness of the antimeltback layer in our 1.55-μm lasers determines to some extent the measured chirp width; for example, lasers with larger antimeltback layer thickness have smaller confinement factor and hence have less chirp. If the laser is modulated at bit rates higher than the relaxation oscillation frequency, the chirp width is significantly enhanced. Since relaxation oscillations are damped in strongly index-guided lasers we expect these lasers to have less chirping at high bit rates (>1 Gb/s) than weakly index-guided lasers. The external cavity-type single-frequency lasers have been fabricated from our multimode lasers.[60] These single-frequency lasers exhibit less chirp (~2) than the multimode lasers due to frequency pulling effect.[61] Such frequency pulling effects can reduce the chirp width of distributed feedback type of lasers also. A chirp of 1Å limits the bit-rate distance product to 160 Gb-km for 1.55-μm transmission systems using conventional silica fibers.

REFERENCES

1. H. Kressel, Ed., *Semiconductor Devices for Optical Communication*, Berlin, New York: Springer-Verlag, 1980, Chap. 9.
2. S. E. Miller and A. G. Chynoweth, Ed., *Optical Fiber Telecommunication*, New York: Academic Press, 1979, Chaps. 1 and 21.
3. Y. Suematsu, "Long-Wavelength Optical Fiber Communication," Proc. IEEE, *71* (June 1983), pp. 692–721.
4. S. R. Nagel, K. L. Walker, and J. B. MacChesney, "Current Status of MCVD: Process and Performance," Tech. Dig. Topical Meeting Opt. Fiber Commun., Phoenix, Ariz., April 1982, Paper TuCC2.
5. R. A. Linke et al., "A 1 Gb/s Transmission Experiment Over 101 km of Single Mode Fiber Using a 1.55 μm Ridge Guide C^3-Laser," Electron. Lett, *20* (1984), pp. 472–4.
6. B. L. Kasper et al., "A 161.5 km Transmission Experiment at 420 Mb/s," Proc. Eur. Conf. Opt. Commun., Stuttgart, August 1984.
7. N. K. Dutta and R. J. Nelson, "The Case for Auger Recombination in InGaAsP," J. Appl. Phys., *53* (January 1982), pp. 74–92.
8. A. Sugimura, "Band to Band Auger Recombination Effect on InGaAsP Laser Threshold," IEEE J. Quantum Electron., *QE-17* (May 1981), pp. 627–35.
9. A. R. Adams et al., "The Temperature Dependence of the Efficiency and Threshold Current of InGaAsP Lasers Related to Intervalence Band Adsorption," Jap. J. Appl. Phys., *19* (October 1980), pp. L621–4.
10. K. Sakai, S. Akiba, and T. Yamamota, "InGaAsP Double Heterostructure Laser," Jap. J. Appl. Phys., *16* (1977), pp. 2043–5.
11. R. J. Nelson and N. K. Dutta, "Self Sustained Pulsations and Negative Resistance Behavior in InGaAsP Double Heterostructure Lasers," Appl. Phys. Lett., *37* (November 1980), pp. 769–71.
12. N. K. Dutta, D. P. Wilt, and R. J. Nelson, "Analysis of Leakage Currents in 1.3 μm InGaAsP Real-Index-Guided-Lasers," J. Lightwave Technol., *LT-2* (June 1984), pp. 201–8.
13. A. R. Beattie and P. T. Landsberg, "Auger Effect in Semiconductors," Proc. Roy. Soc. London, *249* (1959), pp. 16–28.
14. F. Stern, "Calculated Spectral Dependence of Gain in Excited GaAs," J. Appl. Phys., *47* (December 1976), pp. 5382–6.
15. N. K. Dutta, "Gain-Current Relation for InGaAsP Lasers," J. Appl. Phys., *52* (January 1981), pp. 55–60.
16. A. Mozer et al., "Losses in GaInAsP/InP and GaAlSb(As)/GaSb Lasers—The Influence of Split-Off Valence Band," IEEE J. Quantum Electron., *QE-19* (June 1983), pp. 913–6.

17. B. Sermage et al., "Photoexcited Carrier Lifetime and Auger Recombination in 1.3 μm InGaAsP," Appl. Phys. Lett., *42* (February 1983), pp. 259–61.
18. G. H. B. Thompson, "Analysis of Radiative and Nonradiative Recombination Law in Lightly Doped InGaAsP Lasers," Electron. Lett., *19* (March 1983), pp. 154–5.
19. T. Uji, K. Iwamoto, and R. Lang, "Nonradiative Recombination in InGaAsP-InP Light Sources Causing Light Emitting Diode Output Saturation and Strong Laser-Threshold-Current Temperature Sensitivity," Appl. Phys. Lett., *38* (February 1981), pp. 193–5.
20. G. P. Agrawal and N. K. Dutta, "Effect of Auger Recombination on the Threshold Characteristics of Gain Guided InGaAsP Lasers," Electron. Lett., *19* (November 1983), pp. 974–6.
21. C. B. Su et al., "Measurement of Radiative and Auger Recombination Rates in p-Type InGaAsP Diode Lasers," Electron. Lett., *18* (July 1982), pp. 595–6.
22. E. Winter and E. P. Ippen, "Nonlinear Carrier Dynamics in GaInAsP Compounds," Appl. Phys. Lett., *44* (May 1984), pp. 999–1001.
23. R. J. Nelson and N. K. Dutta, "Calculated Auger Rates and Temperature Dependence of Threshold for Semiconductor Lasers Emitting at 1.3 and 1.55 μm," J. Appl. Phys., *54* (June 1983), pp. 2923–9.
24. W. B. Joyce and R. W. Dixon, "Analytic Approximations for the Fermi Energy of an Ideal Fermi Gas," *31* (September 1977), pp. 354–6.
25. E. O. Kane, "Band Structure of Indium Antimoride," J. Phys. Chem. Solids., *1* (1957), pp. 249–61.
26. A. Haug, "Auger Recombination in InGaAsP," Appl. Phys. Lett., *42* (March 1983), pp. 512–4.
27. J. R. Chelikowsky and M. L. Cohen, "Nonlocal Pseudo Potential Calculation for the Electronic Structure of Eleven Diamond and Zinc-Blende Semiconductors," Phys. Rev., B, *14* (July 1976), pp. 556–82.
28. N. K. Dutta et al., "Temperature Dependence of Threshold Current of Injection Lasers for Short Pulse Excitation," Appl. Phys. Lett., *44* (May 1984), pp. 943–4.
29. N. K. Dutta, "Calculated Temperature Dependence of Threshold Current of GaAs-AlGaAs Double Heterostructure Lasers," J. Appl. Phys., *52* (January 1981), pp. 70–3.
30. P. J. Anthony and N. E. Schumaker "Ambipolar Transport in Double Heterostructure Injection Lasers," IEEE Electron Dev. Lett., *EDL-1* (April 1980), pp. 58–60.
31. B. Etienne et al., "Influence of Hot Carriers on Temperature Dependence of Threshold of InGaAsP Lasers," Appl. Phys. Lett., *41* (1982), p. 1018.
32. S. Yamakoshi et al., "Direct Observation of Electron Leakage in InGaAsP/InP Double Heterostructure," Appl. Phys. Lett., *40* (January 1982), pp. 144–6.
33. C. H. Henry et al., "Minority Carrier Lifetime and Luminescence Efficiency of 1.3 μm InGaAsP-InP Double Heterostructure Lasers," IEEE J. Quantum Electron., *QE-19* (June 1983), pp. 905–13.
34. C. H. Henry et al., "The Effect of Intervalence Band Absorption on the Thermal Behavior of InGaAsP Lasers," IEEE J. Quantum Electron., *QE-19* (June 1983), pp. 747–53.
35. J. K. Butler, "Theory of Transverse Cavity Mode Selection in Homojunction and Heterojunction Semiconductor Diode Lasers," J. Appl. Phys., *42* (October 1971), pp. 4447–57.
36. E. J. Flynn and D. A. Ackermann, unpublished work.
37. H. Ishikawa et al., "V-Grooved Substrate Buried Heterostructure InGaAsP/InP Laser With Smooth Far Field Pattern and Stable Aging Characteristics," Jap. J. Appl. Phys., *21*, Suppl. 21–1 (1982), pp. 435–6.
38. M. Hirao et al., "Fabrication and Characterization of Narrow Stripe InGaAsP/InP Buried Heterostructure Lasers," J. Appl. Phys., *51* (August 1980), pp. 4539–40.
39. R. J. Nelson et al., "CW Electrooptical Properties of InGaAsP ($\lambda = 1.3$ μm) Buried Heterostructure Lasers," IEEE J. Quantum Electron., *QE-17* (February 1981), pp. 202–7.
40. I. Mito et al., "InGaAsP Double Channel Planar Buried Heterostructure Lasers Diode (DCPBH-LD) With Effective Current Confinement IEEE," J. Lightwave Technol., *LT-1* (March 1983), pp. 195–202.
41. P. D. Wright et al., "Long Wavelength InGaAsP Lasers," Fiber Optic Communication/Local Area Network Conf., Atlantic City, N.J., October 10–14, 1983.
42. S. Akiba et al., "Low Threshold Current Distributed-Feedback InGaAsP/InP CW Lasers," Electron. Lett., *18* (January 1982), pp. 77–8.
43. N. K. Dutta et al., "InGaAsP Laser With Light T_0," IEEE J. Quantum Electron., *QE-18* (October 1982), pp. 1414–6.
44. N. K. Dutta et al., "Criterion for Improved Linearity of 1.3 μm InGaAsP-InP Buried

Heterostructure Lasers," IEEE J. Lightwave Technol., *LT-2* (April 1984), pp. 160–4.
45. F. R. Nash et al., "Implementation of the Proposed Reliability Assurance Strategy for an InGaAsP/InP, Planar Mesa, Buried Heterostructure Laser Operating at 1.3 μm for Use in a Submarine Cable," AT&T Tech. J., *64* (March 1985), pp. 809–60.
46. D. P. Wilt et al., "Channelled Substrate Buried Heterostructure Laser Using Hybrid VPE and LPE," J. Appl. Phys., *56* (August 1984), pp. 710–2.
47. H. Ishikawa et al., "V-Grooved Substrate Buried Heterostructure InGaAsP/InP Laser by One Step Epitaxy," J. Appl. Phys., *53* (April 1982), pp. 2851–3.
48. D. P. Wilt et al., "Channelled Substrate Laser Using Fe Implantation for Current Confinement," Appl. Phys. Lett., *44* (1984), pp. 290–3.
49. D. P. Wilt et al., unpublished work.
50. N. K. Dutta "Improved Linearity and Kink Criterion for 1.3 μm InGaAsP-InP Channelled Substrate BH Laser," Appl. Phys. Lett., *44* (March 1984), pp. 483–5.
51. P. S. Henry, unpublished work.
52. S. D. Personick, "Receiver Design for Digital Fiber Optic Communication System—Part I," B.S.T.J., *52* (July–August 1973), pp. 843–74.
53. J. E. Midwinter, *Optical Fibers for Transmission*, New York: Wiley, 1979, Chap. 8.
54. W. T. Tsang, N. A. Olsson, and R. A. Logan, "High Speed Direct Single-Frequency Modulation With Large Tuning Rate and Frequency Excursion in Cleaved Coupled Cavity Semiconductor Laser," Appl. Phys. Lett., *42* (March 1983), pp. 650–3.
55. K.-Y. Liou, "Single Longitudial Mode Operation of Injection Laser Coupled to a GRINROD External Cavity," Electron. Lett., *19* (1983), pp. 750–2.
56. N. A. Olsson, N. K. Dutta, and K-Y. Liou, "Dynamic Line Width of Amplitude Modulated Single Longitudinal Mode Semiconductor Lasers," Electron. Lett., *20* (February 1984), pp. 121–2.
57. N. K. Dutta et al., "Frequency Chirp Under Current Modulation in InGaAsP Injection Lasers," J. Appl. Phys., *56* (October 1984), pp. 2167–9.
58. K. Kishiuo, S. Aoki, and Y. Suematsu, "Wavelength Variation of 1.6 μm Wavelength Buried Heterostructure GaInAsP/InP Lasers Due to Direct Modulation," IEEE J. Quantum Electron., *QE-18* (1982), pp. 343–51.
59. F. Bosch, G. L. Dybwad, and C. B. Swan, U.S. Patent 4-317-236, February 1982.
60. N. A. Olsson et al., "2 Gb/s Operation of Single Longitudinal Mode 1.5 μm Double-Channel Planar Buried Heterostructure C^3 Lasers," Electron. Lett., *20* (May 1984), pp. 395–7.
61. G. P. Agrawal, N. A. Olsson, and N. K. Dutta, "Reduced Chirping in Coupled-Cavity Semiconductor Lasers," Appl. Phys. Lett., *45* (July 1984), pp. 119–21.

AUTHORS

P. Besomi, Ph.D. (Materials Science), 1982, Northwestern University; AT&T Bell Laboratories, 1981–1984; General Optronics, 1984—. Mr. Besomi is presently product line manager for InGaAsP LEDs and detectors.

Robert L. Brown, Certificate (Electrical Engineering), 1953, RCA Institute, N.Y.; AT&T Bell Laboratories, 1953—. Mr. Brown has worked on Si, GaAs, and InP devices. He is currently a Member of Technical Staff and is the author of 46 publications.

Richard W. Dixon, A.B. (Engineering and Applied Physics), 1958, Harvard College; M.S. and Ph.D. (Engineering and Applied Physics), Harvard University, in 1960 and 1964, respectively; AT&T Bell Laboratories, 1965—. Mr. Dixon is currently the Director of the Lightwave Devices Laboratory of AT&T Bell Laboratories.

Niloy K. Dutta, B.Sc. (Physics, Honours), 1972, and M.Sc. (Physics), 1974, St. Stephen's College, New Delhi, India; Ph.D. (Physics), 1978, Cornell Uni-

versity, Ithaca, N.Y.; AT&T Bell Laboratories, 1979—. Mr. Dutta was a Postdoctoral Associate at Cornell University for one and half years. During his doctoral and postdoctoral work at Cornell, he worked on spin-flip Raman lasers, infrared spectroscopy, soft X-ray lasers, and nonlinear optics with nonmonochromatic waves. Since 1979 he has been a Member of Technical Staff at AT&T Bell Laboratories, where he has contributed extensively to the research and development of InGaAsP semiconductor lasers. Mr. Dutta has authored or coauthored more than 80 publications on his work and one book on *Long Wavelength Semiconductor Lasers*, which is to be published. Senior Member, IEEE. Member, American Physical Society, Optical Society of America.

Ronald J. Nelson, B.A. (Physics), MacMurray College; M.S. and Ph.D., University of Illinois, Urbana; AT&T Bell Laboratories, 1976–1984; Lytel Inc., 1984—. At AT&T Bell Laboratories Mr. Nelson carried out studies of deep trap states in GaAlAs, interfacial recombination in GaAlAs-GaAs heterojunctions, and both radiative and nonradiative processes in GaAs material and laser devices. He was also involved in the fabrication and characterization of InGaAsP injection lasers. Mr. Nelson was appointed Supervisor of the Long Wavelength Device group in 1979. Member, American Physical Society.

Randall B. Wilson, B.S. (Physical Chemistry), 1975, University of California, Berkeley; Ph.D. (Chemistry), 1979, The Massachusetts Institute of Technology; AT&T Bell Laboratories, 1979–1984; Lytel Inc., 1984—. At AT&T Bell Laboratories Mr. Wilson was involved in studying the properties and LPE growth of III-V materials for long-wavelength device applications.

Daniel P. Wilt, B.A. (Mathematics and Physics), 1976, University of Southern California, University Park; Ph.D. (Applied Physics), 1981, California Institute of Technology, Pasadena; AT&T Bell Laboratories, 1981—. Mr. Wilt is currently pursuing research in the field of semiconductor lasers. He was appointed Supervisor of Lightwave Materials group in 1984. Member, Phi Beta Kappa, and Phi Kappa Phi.

AT&T Technical Journal
Vol. 64, No. 8, October 1985
Printed in U.S.A.

Equalizing Without Altering or Detecting Data

By G. J. FOSCHINI*

(Manuscript received August 27, 1984)

For terrestrial digital radio systems that use Quadrature Amplitude Modulation, the idea of adapting equalizers to multipath distortion, without relying on accurate data estimates, is attractive. Prompt adaptation following a severe fade, when accurate data estimates are unavailable, is useful for reducing outage time. To avoid processing and administrative overhead, the adaptation method should not involve violating the transmitted signal with the insertion of equalizer training signals. We approach this kind of equalization by building on an algorithm of D. Godard (IEEE Transactions on Communications, November 1980)[1] that was devised for voiceband polling networks. The method involves a very simple tap update procedure. However, the technique lacks the foundation of the years of analysis and experimentation that underlie least-mean-square adaptation algorithms. The main purpose of this paper is to present new findings, including (1) a proof that the algorithm, thought to require special equalizer initialization, converges regardless of initialization (this offers useful flexibility in digital radio systems, since, after a severe fade, the algorithm could start with any tap misalignment); (2) a preliminary look at convergence speed suggesting the possibility of significant outage reduction; (3) an algorithm that provides phase coherence (the original algorithm requires a follow-on phase-locked loop); and (4) an algorithm for cross-polarization cancellation as well as equalization.

I. INTRODUCTION

1.1 The problem of prompt data detection

Consider a Quadrature Amplitude Modulation (QAM) digital radio signal (or a dually polarized QAM pair) propagating through a medium

* AT&T Bell Laboratories.

subject to slowly, randomly varying, frequency selective fades (and cross-polarization coupling). On certain occasions the loss of signal can be so complete that a theoretically optimum receiver could not detect the data. Subsequently, a strong signal returns, but carrier and timing may have wandered and the medium may have significantly changed its dispersive character. The objective is to detect the data symbols forthwith as the signal strength returns.

It is the uncertainty about the various features of the received signal, apart from the inherent uncertainty associated with the information symbols (and additive noise), that slows the recapture process. Carrier frequency and phase, and timing frequency and phase are all to some degree uncertain. Moreover, the 2×2 matrix transfer characteristic of the dispersive medium (the diagonal elements describe the co-polarization transfer characteristics and the off-diagonal terms express the couplings between polarizations) is also uncertain. This medium must be equalized to enable accurate data detection. During the bootstrapping, reliable data estimates are unavailable. Consequently, data-directed Minimum-Mean-Square Error (MMSE) equalization is not feasible.

In this paper a method of equalization is analyzed that does not require the availability of data estimates. The method involves tap adjustments based on simple computations using samples of the received QAM signal and of the equalizer output. For simplicity, in the following sections equalization (along with cross-polarization cancellation) is considered in isolation assuming carrier and timing recovery have somehow been accommodated. We stress that equalization is one part of the bootstrapping process. At the time of this writing, carrier recovery is a topic of research. One promising approach employs a quartic nonlinearity. Later we will say more about carrier recovery. Regarding timing frequency, we anticipate that the squared-envelope method is adequate in most applications. Practical realizations of the systems we analyze are assumed to employ a sufficient number of fractionally spaced taps to be quite robust to timing phase.

1.2 Are probing tones not needed?

The approach to equalization that we treat leaves the standard form of the transmitted signal inviolate. (In contrast, one could monitor the medium with real-time measurements and adjust equalizers on the basis of the measurements.) Is the standard (no modification at the transmitter) signal already a media probing signal? Is it also a control signal that arranges for equalizers and cross-polarization cancellors to automatically align in response to reasonable real-time, digital signal processing? A practical affirmative answer would enable one to avoid

the processing and administrative overhead associated with altering the form of the transmitted signal.

1.3 Outline of results

In this paper we take an approach to equalization, without altering or detecting data, that was originally designed for expediting start-up in equalizers in voice band polling networks. The method, originated by D. Godard,[2] involves a very simple tap update procedure and for that reason is especially attractive. [See Ref. 3 for an earlier paper providing a method for PAM (but not QAM) signals. Related research has been conducted.[4-6]] The algorithm assumes the average squared constellation vector is zero and is presented in Section IV. First some background is needed. In Section II the basic model for single polarization transmission is presented. Section III discusses a general view of the tap evolution that will be used.

Godard's algorithm lacks the foundation of the years of analysis and experimentation that underlie least-mean-square adaptation algorithms. The main purpose of this paper is to present new findings on the mathematical theory of this little-understood algorithm.

The algorithm was thought to require special tap initialization to converge. We show that the algorithm converges regardless of initialization (Section V). This flexibility is significant given the vagaries of tap misalignments that could be associated with severe channel fading.

In Section VI we take a preliminary look at the subject of convergence speed. This is accomplished by reinterpreting published numerical work[1] (aimed at voiceband systems) for digital radio applications. While a mathematical analysis of the transient behavior seems intractable, Section VI explains that considerable insight can be obtained from the analysis of a much simpler related problem.

Two new algorithms related to the Godard algorithm are given. The first (Section VII) provides phase coherence as part of the equalization process for hypothetical systems employing highly stable oscillators. The original algorithm required a follow-on phase-locked loop. Section VIII presents an algorithm that provides for cross-polarization cancellation as well as equalization.

Throughout much of this paper the mathematical theory idealizes assuming an infinite tap equalizer. This is because it is very awkward to deal analytically with the finite tap case. One is left wondering if there is any pitfall associated with the finite tap algorithm. The current status of this issue is addressed in Section IX. There is some evidence that the infinite tap—finite tap contrast is nearly as tame as it is with MMSE equalization.

A second purpose of this paper is to answer the question of whether the potential for using the aforementioned algorithm in digital radio

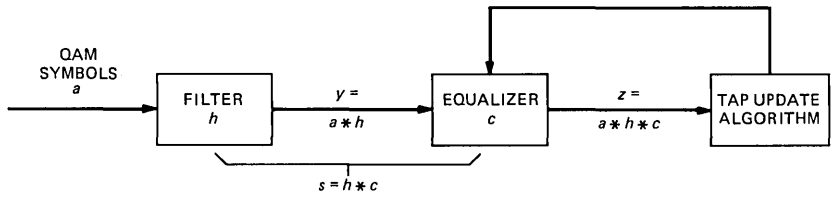


Fig. 1—Simplified baseband model.

systems is significant enough to warrant detailed study. An affirmative answer is given.

A discussion is given in Section X. The Appendix contains information on QAM constellation moments.

II. THE MODEL

Since our primary goal is the presentation of theoretical results, a simple setting is used. We work with the equivalent baseband model shown in Fig. 1. The complex data sequence is denoted a. The elements of $a = (\cdots a_0, a_1, \cdots)$ represent independent identically distributed choices from a QAM constellation, each point of which is equally likely. We normalize so that $E\,|\,a_n\,|^2 = 1$.

The complex sequence h represents samples of the impulse response of the transmitter and medium combination. The sequence c represents the complex equalizer taps. Using the $*$ symbol for convolution, the sampled impulse response of the channel and equalizer in combination is denoted $s = h * c$, the received data is denoted $y = a * h$, and the sequence after the equalizer is $z = a * h * c$. This notation is consistent with the notation of Ref. 1.

Also, h is assumed to have a continuous Fourier transform devoid of spectral nulls. Consequently, h has a convolution inverse h^{-1} satisfying $h * h^{-1} = \bar{0}1\bar{0}$. By $\bar{0}$ we mean an infinite sequence of zeroes, left directed if preceding a number and right directed if following a number. If $\bar{0}$ is written without abutting a number, it means the sequence of zeroes extending from $-\infty$ to $+\infty$.

A more refined model of the terrestrial digital radio environment would include additive white Guassian noise at the input to the receiver. However, the major interest in this paper is in prompt reestablishment of adequate equalization *after* a cataclysmic event during which the data-detection capability was completely lost (so $P_e \rightarrow 1/2$). The situation is that the medium, despite the presence of additive noise, has the potential of providing adequate performance if only the

equalizer could be properly aligned. In such situations the s/n is generally so large that optimal MMSE equalizer including noise effects is only slightly better than the inverse equalizer,[7] which neglects noise. Once the equalization can provide for a P_e in the neighborhood of 0.1, conventional linear MMSE equalization is an option. To be prudent, unless stated otherwise, this option is assumed. So neglecting the noise is not a substantive shortcoming of the analysis. MMSE Decision Feedback (DF) equalization is an alternative option. The bane of DF has been the possibility of the detection process entering a disasterous error propagation mode. With Godard's algorithm as a fallback, the error propagation is no longer disasterous. The theoretical performance of DF is superior to linear equalization.[8-10] Therefore, in a severe fade, DF helps in forestalling the failure of the decision-directed mode. However, in some important digital radio applications the theoretical advantage of DF over linear is marginal.[7]

III. VECTOR FIELD FOR EQUALIZER TAP EVOLUTION

Generally a tap update is a random vector. The subject of "convergence" of a tap update procedure prompts several interesting questions. What is the underlying trend in the tap evolution? If the tap setting is tending to some target region, how long does it take to get there? What is the long-term equilibrium distribution of the tap settings? Recasting the last question less mathematically: Do the tap settings significantly stray from the target region once they get there? These are difficult questions. In this paper we will primarily attack the first question, say a little about the second, and defer the third. The target region can be defined as constituting those settings for which MMSE can commence. We proposed above that MMSE be used when possible. Consequently, for the third question we defer to MMSE theory.

To address the first question, we shall deal with the mathematical abstraction of vector fields in tap space. That is, at each point in tap space there is a vector, that, when added to the current vector of tap settings, points to where the taps should nominally be set in the immediate future. Anticipating that time spans of interest in tap bootstrapping processes involve over 10^4 symbols, it is sometimes useful to approximate and represent tap evolution in continuous rather than discrete time. The vectors are smooth functions of the taps. The relative magnitudes of the vectors relate to the relative speed of change of the taps.

The fields of interest to us are conservative, that is, they are derived by taking the gradient of a potential function ϕ. The gradient depends on the current tap setting and the channel-impulse response. We stress that in implementations the channel response is not known and the tap changes (gradients) are derived from readily accessible random

variables whose mean values are the desired quantities. (This is the so-called stochastic gradient approach. Conventional MMSE utilizes a stochastic gradient approach requiring accurate data estimates.)

The potential functions we will work with are fourth-degree polynomials in the tap gains and their complex conjugates. Exceptional points, where the field vector is a zero vector, are called stationary points. For potential functions such as fourth-degree polynomials the Hessian matrix

$$\mathscr{H} = \left(\frac{\partial^2 \phi}{\partial \bar{c}_l \partial c_m} \right)$$

types the stationary points. (We are using a bar here for complex conjugation.) This typing stems from a power-series expansion about a stationary point and is as follows:

positive semidefinite $\leftrightarrow$ local minimum
negative semidefinite $\leftrightarrow$ local maximum
indefinite $\leftrightarrow$ unstable equilibrium.

In the sequel, when we say that an algorithm "converges" to some set, we mean that the set constitutes the points of stable equilibrium of the corresponding vector field.

The stochastic gradients of $|z_n|^4$ and $|z_n|^2$ will be used later. We record them for reference for even positive powers Q. We obtain (similar to Ref. 1)

$$\frac{\partial |z_n|^Q}{\partial \bar{c}_k} = \frac{\partial}{\partial \bar{c}_k} [(y'c)(\bar{y}'\bar{c})]^{Q/2} = \frac{Q}{2} [(y'c)(\bar{y}'\bar{c})]^{(Q/2)-1} (y'c)\bar{y}_k.$$

The prime denotes transpose and the bar denotes conjugation. Since $(Q/2) - 1$ is a nonnegative integer the computation of the stochastic gradient of $|z_n|^Q$ involves only multiplications of readily accessible quantities. These quantities are the tap settings and the vector of tap outputs. We note that the taps evolve so that the vector c is a function of time and the notation has suppressed that dependence thus far.

IV. DESCRIPTION OF GODARD'S QUARTIC ALGORITHM

In this section we review the algorithm of Ref. 1 for updating equalizer taps, for which, remarkably, data estimates are not required. The algorithm, if properly initialized, is known to converge to a vector of the form $h * c = (\cdots 0, e^{j\theta}, 0, 0, \cdots)$, which is the ideal Nyquist response, except for the presence of an arbitrary phase θ. In other words, the algorithm converges so that z is the same as a except that the constellation needs to be rotated into position. The four-fold ambiguity associated with positioning the constellation is not a prob-

lem since the data is assumed to be differentially encoded at the transmitter.

Let $c_{[n]}$ denote the tap vector at time n. The tap update procedure is based on a gradient minimization of

$$\phi(c_{[n]}) = E(|z_n|^2 - R)^2, \tag{1}$$

where $R = E|a_n|^4/E|a_n|^2 = E|a_n|^4$ is a constant depending only on the signal constellation. Thus the tap update procedure is

$$c_{[n+1]} = c_{[n]} - \lambda \bar{y}_n z_n(|z_n|^2 - R), \tag{2}$$

where λ is the step size.

One way to motivate eq. (1) is to consider

$$\hat{\phi} = E(|z_n|^2 - |a_n|^2)^2 \tag{3}$$

as an error criterion. Overlook, for a moment, that estimates of $|a_n|^2$ are not available at the receiver immediately after a severe fade. $\hat{\phi}$ has some nice features. After all it measures zero when $z_n = a_n$. The systems we are dealing with are nominally linear, so that it seems reasonable to speculate that when $\hat{\phi}$ is zero, then the system is ideally equalized modulo a rotation of the constellation. We must replace $|a_n|^2$ in (3) by a more reasonable quantity. We replace it by a constant chosen to make the two expectations in (1) and (3) have substantial agreement when expressed in more fundamental form in terms of h and c. (See Ref. 1 for details.) This completes the interpretation of the choice of R.

Recalling that $s = h * c$, the potential $\phi(c)$ is expressed as

$$\phi(c) = 2\,(\sum |s_k|^2)^2 - (2 - E|a_n|^4) \sum |s_k|^4 - 2E|a_n|^4 \sum |s_k|^2 + (E|a_n|^4)^2. \tag{4}$$

Let L^2 be the number of constellation points and recall the normalization $E|a_n|^2 = 1$. Regarding $s = h * c$ as the independent variable, we have

$$\phi(s) = 2\,(\sum |s_k|^2)^2 - \left(\frac{3(L^2+1)}{5(L^2-1)}\right) \sum |s_k|^4 - 2\left(\frac{7L^2-13}{5(L^2-1)}\right) \sum |s_k|^2 + \left(\frac{7L^2-13}{5(L^2-1)}\right)^2. \tag{5}$$

The equality (4), comes from straightforward harmonic analysis (it appears in Ref. 1) and equality (5) uses the Appendix.

In the following sections new results are presented on the theory of quartic algorithms.

V. CONVERGENCE OF THE QUARTIC ALGORITHM

In the previous section we mentioned that the algorithm converges if it is properly initialized. Now we will show that such initialization is not needed for convergence. In a certain sense the algorithm will converge regardless of the initialization. More precisely, we show that the only loci of stability of the vector field in tap space are the family of sets

$$E_k = \{c \mid c * h = e^{j\theta}(\bar{0}1_k\bar{0}), \quad \theta \text{ an arbitrary real}\}$$

Each E_k is an ellipse in tap space since it is expressible as a linear transformation of a circle. By 1_k we mean that 1 occurs in the kth position. The set $E \triangleq \cup_{-\infty}^{\infty} E_k$ are the points of global minima. These points are the ideal Nyquist responses (modulo a phase adjustment of the constellation).

A gradient search can only terminate on one of these circles of local minima. There are no spurious local minima. The only other stationary points are $c = \bar{0}$, which is a local maximum corresponding to the shut-down of the receiver and some points of unstable equilibrium "saddle points." We now substantiate these claims. The mathematical approach used in the remainder of this section is similar to that used in Sections VII and VIII. However, the demonstration here is much simpler.

5.1 Stationary points relative to overall system response

To demonstrate the character of the stationary points, we first discuss the stationary behavior relative to s, from which the nature of the stationary points relative to c will follow.

With respect to the conjugate coordinates $\bar{s}_k$, we take the partial derivative of ϕ and equate to zero to get

$$\left\{\frac{\partial\phi}{\partial\bar{s}_k} = 4\left(\sum_i |s_i|^2\right)s_k - \frac{6}{5}\frac{(L^2+1)}{(L^2-1)}|s_k|^2 s_k - 2\left(\frac{7L^2-13}{5(L^2-1)}\right)s_k = 0\right\}_{k=-\infty}^{\infty}.$$

So $s = \bar{0}$ is a solution. Dividing through by s_k for $s_k \neq 0$ gives a very simple equation for the stationary points. In general, the stationary points are the vectors having the property that there are a finite number $M \geq 0$ of nonzero coordinates all M of which are $(7L^2 - 13)[10M(L^2-1) - 3(L^2+1)]^{-1}$ in squared modulous. These stationary points are typed as

$M = 0$: $s = \bar{0}$, a local maximum
$M = 1$: global minima
$M \geq 2$: unstable equilibria.

The reasoning behind this typing of each case will now be given. Keep in mind the expression of $\phi(s)$ given in eq. (4).

The case $M = 0$.

This corresponds to $s = \bar{0}$, which is a local maximum. Simply note that sufficiently small perturbations of $\bar{0}$ serve to reduce the value of $\phi(s)$. The reason for this reduction is that, for a very small perturbation, the quartic effect is negligible relative to the quadratic one.

The case $M = 1$.

These points are loci of global minima. The global minimum cannot be attained if more than one component of s is nonzero. Indeed, if $\hat{s}$ has more than one nonzero component then $\tilde{s} \triangleq \bar{0}, (\sum |\hat{s}_i|^2)^{1/2}, \bar{0}$ gives $\phi(\tilde{s}) < \phi(\hat{s})$. Say that the ith component of s is the only nonzero component. The function $\phi(s)$ is then minimized when $|s_i|^2 = 1$.

The case $M \geq 2$.

These points are unstable equilibria ("saddles"). The instability for a stationary point s with $M \geq 2$ is shown by first decreasing $\phi(s)$ by a perturbation of two nonzero components of s that leaves $\sum |s_i|^2$ invariant. Secondly, $\phi(s)$ is increased by a perturbation that simply increases the magnitude of a zero component by a sufficiently small positive number.

5.2 Stationary points relative to tap weights

The results of Section 5.1 are only of incidental value since we are interested in the vector field relative to c not relative to s. However, the results thus far can be interpreted to provide what we need, as we now explain.

The operation of convolution of c with h represents an invertible continuous function on tap space. Since $\phi(s)$ is a continuous real function of s, $\phi(s) = \phi(h * c)$ is also a continuous real function of c. Recall the very elementary fact that notions like local maximum, local minimum, and unstable equilibrium are defined as neighborhood properties in tap space. It is a property of continuity that the mappings $h*$ and $h^{-1}*$ leave invariant the entire system of neighborhoods. It is immediate from the results of Section 5.1 that the points in E are the only points of stability in tap space when the taps evolve in accordance with the specified vector field.

For those uncomfortable with the above argument we give another level of detail. Say $\hat{s} = h * \hat{c}$ is a point of local minimum of ϕ. This means that there is an open set in tap space containing $\hat{s}$, where $\phi(\hat{s})$ is the least number achieved. By convolution of elements of this open

set with h, one creates an open set about $\hat{c}$ on which $\phi(h * \hat{c})$ is the least number achieved. Local maxima and unstable equilibria are handled similarly.

VI. A PRELIMINARY LOOK AT CONVERGENCE SPEED

6.1 Interpretation of a published simulation

To consider quantitatively the subject of convergence speed, we fix on a hypothetical example. Namely, we focus on a system with a transmission speed measured in tens of megabauds. Outage time accumulates when $P_e \geq 10^{-3}$, and is assumed to be limited to 150 seconds per year per hop (in a nominal system). The hypothetical channel is assumed to fade in accordance with the Rummler model.[11]

Let $P'(t)$ denote the probability of bit error that an *ideally* equalized system can provide at time t. Then $P'(t) \leq P(t)$, where $P(t)$ is the probability of bit error that is actually achieved at time t. For the purpose of discussion, assume that the clear air s/n is such that if $P(t)$ were identically equal to $P'(t)$, the outage objective would be roundly met. Suppose a fade is so severe that decision direction of the equalization process becomes impossible. When the fade subsides to the point where $P' \leq 10^{-3}$, the objective is to boot (or be booting) the equalizer so that $P \leq 10^{-3}$ occurs with a negligible time lag. An aggressive booting procedure, operating with an uncertain frequency-selective transfer characteristic, is viewed as a key element in achieving substantial outage reduction. The alternative of waiting for the dispersion to clear to the extent that a crude equalizer will open the eye is manifestly unacceptable.

The analysis of the booting process is difficult for two reasons. First, the extremal statistics of fade time dynamics are not well established. Second, even if such a model were available, it would be difficult to mathematically represent the time dynamics of an equalizer based on the Godard algorithm. However, if we assume that the time interval during which $P' > 10^{-3}$ lasts for a few seconds, then an equalizer booting time measured in tens of milliseconds would be negligible in terms of contributing to outage time. Indeed, even a 100-millisecond boot time would be negligible if the preponderance of the time occurs before the level $P' = 10^{-3}$ is down-crossed. Assuming that the channel transfer characteristic does not change appreciably during booting, we take a preliminary look at whether the quartic algorithm can boot an equalizer in tens of milliseconds.

Reference 2 reports simulations of transient responses from a cold start. The examples are for a voiceband application; however, they can be interpreted for a digital radio context. One of the examples is particularly interesting. (See Fig. 7d in Ref. 1.) Although 64 QAM is

not treated, rectangular constellations with 16 and 32 points are considered. The effect on the transient response of increasing from 16 to 32 is negligible. (This is expected because the constellation moments for $L^2 = 16$ points are already nearing their $L \to \infty$ asymptotes, as shown in the Appendix.) The channel used to generate the transient responses has as severe a dispersion as can be expected in the digital radio application. This statement is based on a rough indicator of dispersion, namely $\max_\omega |H(\omega)|^2 / \min_\omega |H(\omega)|^2$. ($H(\omega)$ is the channel transfer characteristic.) The unimodal $|H(\omega)|^2$ had a higher indicator of dispersion than the most extreme of the 25,000 fades in a comprehensive library[7] generated from the Rummler model.[11]

The transient responses of Ref. 2 suggest that booting times of less than 10^5 symbols may be possible. In the digital radio application at tens of megabauds, 10^5 symbols are received in a few milliseconds.

A transient analysis of the algorithm is certainly an ambitious undertaking. However, it is possible to conduct a mathematical analysis of eq. (2) for a hypothetical, real, one-dimensional case. We sketch this analysis in the following subsection. By projecting the convergence behavior of interest into a simple, understandable context, useful insight is gained.

6.2 Analysis of a related one-dimensional evolution

In the much simpler domain of least-mean-square adaptation algorithms,[12] the transient behavior for the one-dimensional case is analyzed. Then a heuristic argument is made to extend the transient analysis to the higher-dimensional case. In the one-dimensional analysis of the Godard algorithm that follows, we will see how the corresponding MSE behavior of the algorithm compares with an MMSE evolution. The MMSE evolution assumes known data at the receiver.

In this one-dimensional case, both h and c are real scalars but the data is complex. In this subsection, a coordinate index is unnecessary so we write c_i for $c_{[i]}$. The gradient algorithm is

$$c_{i+1} = c_i - \lambda \frac{\partial}{\partial c_i} (|z_i|^2 - |R|)^2. \tag{6}$$

For expositional simplicity in the sequel, we work with only the asymptotic form ($L \to \infty$) of the constellation moments. (Section V served to demonstrate that dealing with finite L is not an essential complication. The Appendix shows the moment asymptotes are rapidly approached.) Compute for later use

$$\mu(c_i) \triangleq E(c_{i+1} - c_i \,|\, c_i) = -5.6\lambda h[(hc_i)^3 - hc_i] \tag{7a}$$

$$\sigma^2(c_i)^2 \triangleq E[(c_{i+1} - c_i)^2 \,|\, c_i]$$

$$= \lambda^2 h^2 [68.3(hc_i)^6 - 104(hc_i)^4 + 44.2(hc_i)^2]. \tag{7b}$$

Note that both μ and σ are linear in λh.

Were it not for the stochastic aspect, the evolution would be described by the nonlinear difference equation

$$c_{i+1} = c_i - 5.6\lambda h((hc_i)^3 - hc_i). \tag{8}$$

A crude representation of the evolution (8) with λ small uses a deterministic differential equation. A more refined representation of (6) uses a stochastic differential equation. We look at both of these.

The deterministic differential equation is

$$\frac{dc}{dt} = -5.6\lambda h[(hc)^3 - (hc)], \tag{9a}$$

in the time scale where one unit equals one symbol time. The time, T, that it takes for c to evolve from $c_o \neq 0$ to a target setting c_f is easily seen to be

$$T = \frac{1}{11.2\lambda h^2} \ln\left(\frac{1 - (hc_o)^{-2}}{1 - (hc_f)^{-2}}\right). \tag{9b}$$

The corresponding formulae for the MMSE algorithm are

$$\frac{dc}{dt} = -2\lambda h(hc - 1) \tag{9c}$$

$$T = \frac{1}{2\lambda h^2} \ln\left(\frac{1 - hc_f}{1 - hc_o}\right). \tag{9d}$$

The formula (9d) is consistent with the statement in Ref. 12 that the time constant for the MMSE stochastic-gradient algorithm is $(2\lambda h^2)^{-1}$.

If one uses values in the right-hand side of (9b) and (9d) that are reflective of the digital radio application, the logarithmic term can be shown to be of little consequence in assessing order-of-magnitude effects. The latitude in being able to set c_f so that hc_f is only approximately 1 is crucial to taming the singularity. Reference 12 on MMSE explains that, in higher dimensions, h^2 is replaced by $(trace\ R/N)$, where R is the channel autocorrelation matrix and N is the number of equalizer taps. In data communication applications, a normalized form of the channel is often appropriate to account for AGC. In a one-dimensional context, AGC gives $hc = 1$, trivializing the equalization.

Refining (9a) to bring in the stochastic element of the evolution, we have

$$dc = \mu(c)dt + \sigma(c)d\beta(t), \tag{10}$$

where $(d\beta)/(dt)$ is a standard white-noise process. It is difficult to give a complete analysis of this equation; however, some results are possible.

A mathematical "experiment" was made to test the restoring action of the dynamic represented by (10). The mathematical tools in Ref. 13 were used. Set the tap c at a very large value c_o and consider the transit time to another large value γ. Assume $c_o \gg \gamma \gg c_f$. Under the deterministic evolution, (9a), the expected time until γ is hit for the first time is proportional to γ^{-2}. With the stochastic dynamic, the expected hitting time is proportional to γ^{-4}.

As the tap setting c nears the target region, the first-order term in the power series for $\mu(c)$ and $\sigma^2(c)$ is linear and constant, respectively. This limiting evolution is that of the elastically bound particle (also called Orstein-Uhlenbeck[14]). A complete transient analysis of the dynamics of the elastically bound particle is classical.[14] We have

$$E[hc(t) - 1 \mid c(0) = c_o] = (hc_o - 1)e^{-11.2\lambda h^2 t} \tag{11a}$$

$$\mathrm{Var}[hc(t) - 1 \mid c(0) = c_o] = 0.381\lambda h^2(1 - e^{-22.4\lambda h^2 t}). \tag{11b}$$

We are assuming the quartic algorithm is only for bootstrapping, so $0.381\lambda h^2$ does not correspond to the equilibrium variance.

It is interesting to contrast with MMSE, for which the expected time from start to target is available in closed form using the method of Ref. 13. We get

$$T = \frac{1}{\lambda h^2 + 1.4\lambda^2 h^4}\left[\frac{1 - \left(\frac{1 - hc_f}{1 - hc_o}\right)^{1+\frac{5}{7\lambda h^2}}}{1 + \frac{5}{7\lambda h^2}} + \frac{1}{2}\ln\left(\frac{1 - hc_f}{1 - hc_o}\right)\right]$$

$$\sim \frac{1}{2\lambda h^2}\ln\left(\frac{1 - hc_f}{1 - hc_o}\right) \ (\lambda \text{ small}).$$

Also,

$$E(hc(t) - 1 \mid c(0) = c_o) = (hc_o - 1)e^{-2\lambda h^2 t} \quad \text{for each } t, \tag{11c}$$

and

$$\mathrm{Var}(hc(t) - 1 \mid c(0) = c_o) \sim (hc_o - 1)^2 e^{-1.2\lambda h^2 t} \quad \text{as } t \to \infty. \tag{11d}$$

Noise and quantization effects, which are not included, will predominate for large t and thus serve to bound the MMSE variance above zero.

Interestingly, for both the MMSE and quartic algorithm, the exponential decay that is observed in the simple deterministic analysis is maintained when stochastic effects are included. Evolution under the quartic potential compares favorably with that of the quadratic. However, the apparent advantage of the quartic of a factor of 5.6 in time constant is illusory, as we next discuss.

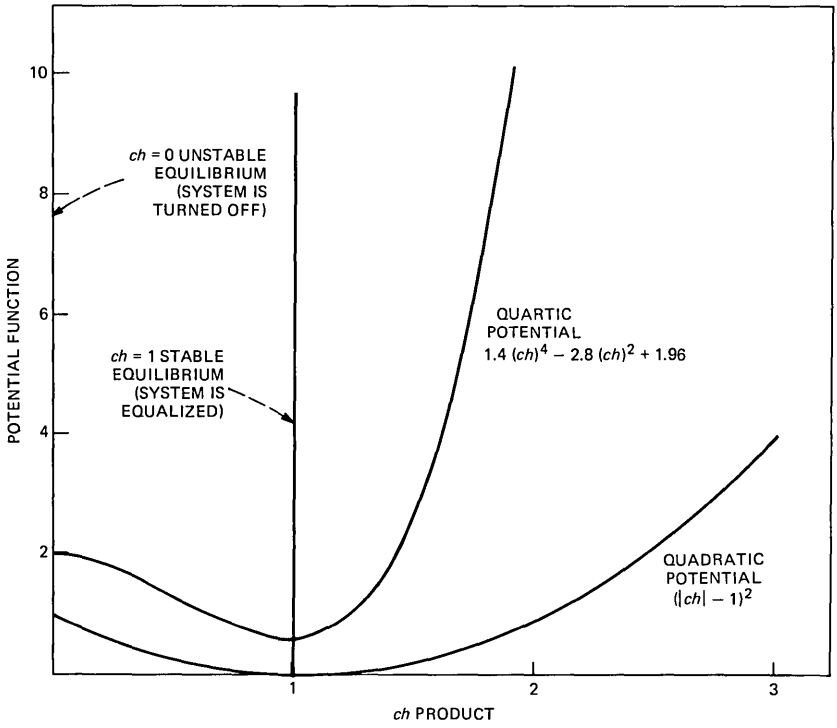


Fig. 2—Potential functions for simple one-dimensional case.

Relation (8a) suggests that one should make λ large to reduce the transit time. However, λ must be set carefully. The differential equation loses its effectiveness as an approximation to (8) once λ gets too large. An analysis of (8) points out that, for λ sufficiently large,

$$|c_{i+1}| > |c_i|,$$

which is a disastrous instability. The wall of the potential well for Godard's algorithm is quartic while, for MMSE, it is quadratic. (See Fig. 2 to contrast the quartic and quadratic potentials.) Consequently, the quartic algorithm requires a smaller λ to avoid instability than MMSE requires. Indeed, in the simulations of Ref. 2 it was noticed that a smaller value of λ was needed. As c_i approaches the target region, one would like λ small to encourage stepping *into* rather than over the target region. A key area of future study is to determine a *dynamic* procedure for prudent setting of λ.

6.3 A heuristic discussion of transient behavior

The results of the previous subsection require refinement to obtain a more global description of the behavior of the trajectories of (6).

More importantly, the vector algorithm needs to be analyzed and that appears to be formidable. These are items for future research. At present we are limited to drawing on what we have developed in Sections V and 6.2.1 to express our current intuition of the qualitative nature of tap evolution: *Far from the target region, there are saddle points representing loci where more than one target region is equally accessible. The evolution is not stymied as the strong random effect forces a choice. Still far from the target region, there is a very strong trend toward the target. The motion slows as the target is approached and becomes that of a particle elastically bound to the elliptical bull's-eye corresponding to the Nyquist responses. Close to the target, the radial motion is not essentially different from an Orstein-Uhlenbeck evolution, except in one respect. The evolution has a tangential indifference that is of no consequence since phase coherence is left for an auxiliary process. The $\overline{0}$ tap setting is an unstable equilibrium that is to be avoided (and that is not difficult). To the extent that the vector tap does not begin too close to zero, the transit time to target would seem to be dominated by the Orstein-Uhlenbeck-like motion.*

VII. REMOVING PHASE AMBIGUITY

A useful feature of the quartic algorithm that we have been discussing is that equalizer convergence does not need carrier recovery. As explained in Ref. 2, the tracking of carrier phase can be carried out using a decision-directed phase-locked loop that will converge once equalization has taken place. For digital radio applications, this feature implies a significant immunity of the equalization process to frequency offset and phase jitter.

On the other hand, if a digital radio system were designed with highly stable oscillators, could we employ a quartic algorithm providing for coherent recovery without a phase-locked loop? After all, standard, decision-directed MMSE equalization provides, upon convergence, for coherent recovery of the signal constellation. This result holds, in principle, under the idealized assumption that the channel is linear and time invariant. In practice, depending on the degree of phase jitter and frequency offset, a phase-locked loop may or may not be required. In this section, we seek to provide an analogous result for nondecision-directed equalization. Specifically, we demonstrate the existence of a potential function that provides for phase recovery as part of equalization. Of course, a QAM constellation is left invariant by 90 degree rotations, so the more precise meaning of the term "phase recovery" is phase recovery modulo 90 degrees.

For expositional simplicity, we develop the potential function for the asymptote of a QAM constellation with infinitely many points. The approach to defining corresponding potentials for finite constel-

lations is the same as for the asymptotic case. The limiting form, as $L \to \infty$, of the potential function given in equation (5) is

$$\phi(s) = 2\left(\sum_i |s_i|^2\right)^2 - 0.6 \sum_i |s_i|^4 - 2.8 \sum |s_i|^2 + 1.96.$$

We will show how to modify $\phi(s)$, but first, for a backdrop, we discuss a much simpler situation in which phase recovery is obtained.

7.1 Orienting a rotated but otherwise perfect constellation

We have seen that the stochastic gradients with respect to c of $E|z_n|^4$ and $E|z_n|^2$ are readily accessible. From (1) and the ensuing discussion, we have that for appropriate choice of A and B, tap settings evolving in accordance with the gradient field of $AE|z_n|^4 + BE|z_n|^2$ converge to an optimal equalizer up to a rotation. Obviously, Re Ez_n^4 also has a readily accessible stochastic gradient. Moreover, Re Ez_n^4 may be useful for phase recovery, as we next indicate.

It is easy to see that Ea_n^4 is a negative number. Let $a_n' = e^{j\theta}a_n$, where θ is an arbitrary phase displacement that does not depend on n. It follows that

$$\eta(\psi) = E \operatorname{Re}(e^{j\psi}a_n')^4$$

is minimized when $\psi = -\theta(\text{mod } \pi/2)$. These are the only minima, whereas $\psi + \theta = \pi/4 \text{ (mod } \pi/2)$ are the only maxima. Consequently, a tap rotating according to the gradient field of $\eta(\psi)$ comes to a stable state when the constellation is correctly oriented.

Based on what we have discussed thus far, one might suspect the existence of a potential of the form $E(A'|z_n|^4 + B'|z_n|^2 + C' \operatorname{Re} z_n^4)$ whose only points of stability are ideal Nyquist responses with recovered carrier. Such potentials exist, as we now show.

7.2 Equalization with orientation

In what follows, positive parameters ν and μ are introduced in the potential function, $\phi(s)$, as follows:

$$\phi(s) = 2\left(\sum |s_k|^2\right)^2 - 2.8 \sum |s_k|^2 - \mu \sum |s_k|^4 - 2\nu \operatorname{Re} \sum s_k^4.$$

Later we will see that ν and μ can be set to get the tap evolution desired. Namely, we can obtain an evolution whose only points of stability are of the form $\overline{0}\epsilon\overline{0}$, $(\epsilon^4 = 1)$.

The gradient of $\phi(s)$ with respect to the conjugate coordinates is

$$\nabla\phi_{\bar{s}_i} = 4 \sum |s_k|^2 s_i - 2.8 s_i - 2\mu |s_i|^2 s_i - 4\nu \bar{s}_i^3. \tag{12}$$

The solutions are the stationary points. The Hessian matrix is denoted $\mathscr{H} = (H_{ij})$ (where $H_{ij} = (\partial^2\phi)/(\partial\bar{s}_i\partial s_j)$). We have

$$H_{ij} = 4 \sum_k |s_k|^2 + 4|s_i|^2 - 2.8 - 4\mu |s_i|^2 \quad \text{for} \quad i = j$$

$$= 4\bar{s}_j s_i \quad \text{for} \quad i \neq j. \tag{13}$$

As in Section 5.2, $\bar{0}$ is easily seen to be a local maximum. Expressing s_i as $|s_i| e^{j\theta_i}$, eq. (12) for the non-null stationary points becomes

$$-4\nu |s_i|^2 e^{-4j\theta_i} - 2\mu |s_i|^2 - 2.8 + 4 \sum |s_k|^2 = 0$$

$$\text{or} \quad |s_i|^2 = \frac{4 \sum_k |s_k|^2 - 2.8}{4\nu e^{-4j\theta_i} + 2\mu}. \tag{14}$$

From the nonnegativity of $|s_i|^2$, we conclude that

$$e^{-4j\theta_i} = \pm 1.$$

But $e^{-4j\theta_i} = -1$ cannot be associated with a local minimum. Just look at (14) and notice $e^{-4j\theta_i} = e^{+4j\theta_i} = -1$ implies Re $\sum_k s_k^4 > 0$ so the mapping $s \to \cdots s_{i-1}, s_i e^{j\zeta}, s_{i+1} \cdots$ reduces the value of ϕ. The local minima of ϕ must have $e^{j4\theta_i} = +1$. These minima satisfy

$$|s_i|^2 = \frac{4 \sum |s_\mu|^2 - 2.8}{4\nu + 2\mu}. \tag{14a}$$

Each minimum has all nonzero coordinates of equal modulus, all equal to the right-hand side of (14). Say there are M nonzero coordinates; then

$$|s_i|^2 = \frac{1.4}{2M - 2\nu - \mu}.$$

To ascertain whether the stationary points are local minima, or points of instability (saddle points), we need to determine the nature of the Hessian at the stationary points. To effectively deal with the Hessian, it is mathematically convenient to permute coordinates so that the coordinates 1 through M are the ones for which (14a) holds. Some notation is also needed. $0_{p,q}$ represents a matrix with p rows and q columns in which each element is zero. I_p denotes a $p \times p$ identity matrix. The Hessian is

$$\mathscr{H} = \frac{2.8}{2M - \mu - 2\nu} \left\{ \begin{bmatrix} 0_{\infty,\infty} & 0_{\infty,M} & 0_{\infty,\infty} \\ 0_{M,\infty} & (2\nu - \mu) I_M & 0_{M,\infty} \\ 0_{\infty,\infty} & 0_{\infty,M} & 2(\nu + \mu) I_\infty \end{bmatrix} + 2s\bar{s}' \right\}. \tag{15}$$

Notice $\mathscr{H}$ is expressed as the sum of a diagonal matrix and a dyad. The nonzero elements of s in eq. (15) are all ± 1 or $\pm j$. The special structure of $\mathscr{H}$ allows the spectrum of $\mathscr{H}$ to be easily found. The nonzero eigenvalues are

$$\frac{0.7(2 + 2\nu - \mu)}{2M - 2\mu - 2} \text{ of multiplicity one}$$

$$\frac{0.7(2\nu - \mu)}{2M - 2\mu - 2} \text{ of multiplicity } M - 1$$

$$\frac{0.7(\mu + 2\nu)}{2M - 2\mu - 2\nu} \text{ of infinite multiplicity.}$$

It remains now to choose μ and ν so that the only stable equilibria are of the form $\bar{0}\epsilon\bar{0}$, where $\epsilon^4 = 1$.

For $|s_i| = 1$ when $M = 1$ choose $\mu + 2\nu = 0.6$. From the spectrum of eigenvalues it follows that if we have

$$2 - \mu + 2\nu > 0 \quad -\mu + 2\nu < 0,$$

then $M = 1$ gives a positive semidefinite Hessian and $M > 1$ gives an indefinite Hessian. For example, $\mu = 0.45$ and $\nu = 0.075$ satisfies all requirements. So the only stable equilibria of $\phi(s)$ with $\mu = 0.45$ and $\nu = 0.075$ are of the form $\bar{0}\epsilon\bar{0}$ with $\epsilon^4 = 1$.

Again, as in Section 5.2.1, we have described the stationary behavior of the gradient field with respect to s and not with respect to c. The argument extends to tap space in the same manner as in Section 5.2.2.

VIII. ALGORITHM FOR CROSS-POLARIZATION CANCELLATION AS WELL AS EQUALIZATION

In this section we develop the theory for a cross-polarization cancellation algorithm. We will establish that a 2×2 matrix equalizer will converge so that both receiver outputs are free of Intersymbol Interference (ISI) and cross-polarization interference. There is a possibility that, despite the perfect decoupling, one or both polarizations may be transposed. The taps evolve in accordance with the gradient of a vector potential that will be introduced shortly. Upon convergence, phase needs to be recovered by a pair of phase-locked loops. This transposition ambiguity is easily resolved in practice. For example, the polarizations may be "tagged" by the scrambling process. When necessary, the procedure can be reinitialized to attempt to avoid locking onto the undesired polarization.

Some notation needs to be introduced. We need a two-dimensional setting to account for horizontally and vertically polarized signals. Here c and h are 2×2 matrices. The vectors (z_H, z_V) and (a, b) are related as follows:

$$\begin{pmatrix} z_H \\ z_V \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} * \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix} * \begin{pmatrix} a \\ b \end{pmatrix}.$$

The individual elements of z_H, z_V, a and b are denoted by subscripts. We use s to denote the matrix

$$c * h = \begin{pmatrix} c_{11} * h_{11} + c_{12} * h_{21} & c_{11} * h_{12} + c_{12} * h_{22} \\ c_{21} * h_{11} + c_{22} * h_{21} & c_{22} * h_{21} + c_{22} * h_{22} \end{pmatrix}.$$

The matrix h is assumed to be nonsingular so that h^{-1} exists

$$\left(h^{-1} * h = \begin{pmatrix} \overline{0}, 1_0, \overline{0} & \\ \overline{0} & \overline{0}, 1_0, \overline{0} \end{pmatrix} \right).$$

The components of the vector (a, b) represent the QAM data sequence driving the horizontal and vertical polarizations. Of course, the elements of a and b are all independent.

We employ a vector criterion; specifically we want

$$\begin{pmatrix} \min\limits_{(c_{11},c_{12})} E(|z_{Hn}|^2 - E|a_n|^4)^2 \\ \min\limits_{(c_{21},c_{22})} E(|z_{Vn}|^2 - E|b_n|^4)^2 \end{pmatrix}.$$

The advisability of this vector criterion will become apparent. Notice optimization of these two components proceed independently of each other in that the first component involves c_{11} and c_{12} while the second involves c_{21} and c_{22}.

We next show how to express these two expectations in terms of the components of s (denoted $s_{ij}(k)i, j = 1, 2$) and the moments of a_n and b_n. By symmetry it will be enough to make this demonstration for the first expectation. Again we normalize $E|a_n|^2 = E|b_n|^2 = 1$. Since

$$z_{Hn} = \sum_k (s_{11}(k)a_{n-k} + s_{12}(k)b_{n-k}),$$

we obtain

$$E|z_{Hn}|^4 = E\left\{ \sum_k \sum_l \sum_p \sum_q (s_{11}(k)a_{n-k} + s_{12}(k)b_{n-k})(\bar{s}_{11}(l)\bar{a}_{n-l} + \bar{s}_{12}(l)\bar{b}_{n-l}) \times (s_{11}(p)a_{n-p} + s_{12}(p)b_{n-p})(\bar{s}_{11}(q)\bar{a}_{n-q} + \bar{s}_{12}(q)\bar{b}_{n-q}) \right\}.$$

Consider the product inside the quadruple sum. Of the 16 terms only the six involving $a_{n-k}\bar{a}_{n-l}a_{n-p}\bar{a}_{n-q}$, $a_{n-k}\bar{a}_{n-l}b_{n-p}\bar{b}_{n-q}$, $a_{n-k}\bar{b}_{n-l}b_{n-p}\bar{a}_{n-q}$, $b_{n-k}\bar{b}_{n-l}b_{n-p}\bar{b}_{n-q}$, $b_{n-k}\bar{b}_{n-l}a_{n-p}\bar{a}_{n-q}$ and $b_{n-k}\bar{a}_{n-k}a_{n-p}\bar{b}_{n-q}$ give nonzero expectations. This simplification follows by recalling that a and b are independent and $Ea^j = Eb^j = \overline{0}$ for $j = 1, 2, 3$. Concerning the six terms, we note that the last three terms become the same as the first three if we transpose a and b. Moreover, the second and third terms are exactly the same since the only apparent differences are in the labelings of indices that are being summed. So we need to ascertain

only the first two of the six expectations and symmetry will give the rest. The two sums are already available. Namely,

$$\sum_k \sum_l \sum_p \sum_q E(s_{11}(k)a_{n-k}\bar{s}_{11}(l)\bar{a}_{n-1}s_{11}(p)a_{n-p}\bar{s}_{11}(q)\bar{a}_{n-q})$$

$$= (E|a_n|^4 - 2)\sum_k |s_{11}(k)|^4 + 2\left(\sum |s_{11}(k)|^2\right)^2 \quad (16)$$

and

$$\sum_k \sum_l \sum_p \sum_q E(s_{11}(k)a_{n-k}\bar{s}_{11}(l)\bar{a}_{n-l}s_{12}(p)b_{n-p}\bar{s}_{12}(q)\bar{b}_{n-q})$$

$$= \left(\sum_k |s_{11}(k)|^2\right)\left(\sum_k |s_{12}(k)|^2\right). \quad (17)$$

Using the aforementioned symmetry gives, for the six terms comprising $E|z_{Hn}|^4$, four copies of (17), and in addition to (16), its counterpart with s_{11} and s_{12} interchanged. To compute $E(|z_{Hn}|^2 - E|a_n|^4)^2$, we also need $E|z_{Hn}|^2 = E|a_n|^2(\sum|s_{11}(k)|^2 + \sum|s_{12}(k)|^2)$. At this point we can substitute all the terms making up $E(|z_{Hn}|^2 - E|a_n|^4)^2$ and rearrange to get

$$E(|z_{Hn}|^2 - E|a_n|^4)^2 = 2(\sum|s_{11}(k)|^2 + \sum|s_{12}(k)|^2)^2$$
$$+ (E|a_n|^4 - 2)(\sum|s_{11}(k)|^4 + |s_{12}(k)|^4)$$
$$-2E|a_n|^4(\sum|s_{11}(k)|^2 + \sum|s_{12}(k)|^2) + (E|a_n|^4)^2. \quad (18)$$

We introduce a new sequence $S(k)$ obtained by alternating $s_{11}(k)$ and $s_{12}(k)$ elements. Thus $S(0) = s_{11}(0)$, $S(1) = s_{12}(0)$, $S(2) = s_{11}(1)$, $S(3) = s_{12}(1)$, $\cdots$ and $S(-1) = s_{12}(-1)$, $S(-2) = s_{11}(-1)$, $S(-2) = s_{12}(-2)$, $S(-3) = s_{11}(-2)\cdots$. Then (18) becomes the same as (5) with S replacing s. The invertibility of h enables us to achieve the optimum value.

The stationary behavior of (18) now follows immediately from the results in Section 5.1. The minimizing c_{11} and c_{12} satisfy

$$\begin{pmatrix} c_{11} * h_{11} + c_{12} * h_{21} \\ c_{11} * h_{12} + c_{12} * h_{22} \end{pmatrix} = \begin{pmatrix} \bar{0} \\ \bar{0}e^{j\theta}\bar{0} \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} \bar{0}e^{j\theta}\bar{0} \\ \bar{0} \end{pmatrix}.$$

We are now in a predicament analogous to that at the end of Section 5.1. As the results stand, they apply to s, not to c. A very straightforward vector matrix extension of the argument in Section 5.1 gives the results desired for c.

IX. FINITE NUMBER OF TAPS

Thus far, the theory has been idealized in the sense that infinitely many taps were assumed. Can we be sure that, by using a large number of taps, an equalizer will behave in essentially the same way as an

infinite tap equalizer? Unlike MMSE theory, the theory of the quartic algorithm with finitely many taps is awkward to treat analytically. Partial, positive results are available and are conveyed here.

The next subsection enhances the quartic algorithm to alleviate one of the difficulties arising with a finite equalizer. The following subsection mentions very special related examples, where, with a finite number of taps, the convergence theory is complete and satisfactory. The next subsection reviews the status of the finite tap issue.

The related topic of quantifying the *number* of taps needed in digital radio applications seems best addressed by computer-aided analysis (as with MMSE equalization[15]). We emphasize that, in the digital radio application, we are *not* aiming to equalize pathological $H(\omega)$ with severe in-band nulls (and consequently an unreasonable number of taps to approximate h^{-1}). Rather, as indicated in Section VI, the interest is in equalizing when $H(\omega)$ can support a P_e' of the order of 10^{-2}. For implementations, one would expect to use fractionally spaced rather than synchronously spaced taps. A numerical example is reported in Section 9.4.

9.1 Centering the tap weight distribution

The feature that the algorithm has no preference as to which tap should be the reference tap implies that, with finitely many taps, the tap weight distribution could crowd to one end of the equalizer. To avoid a lopsided tap-weight distribution one could periodically (say every few hundred symbols) have computed the center of gravity of the tap weights and then shift the weights to situate the balance point as close as possible to the center tap. When the quartic algorithm is in use, outage time is registering, so no additional outage is caused by shifting. The centering helps approximate the effect of infinitely many taps that the theory has required. A computationally simpler alternative to the center of gravity method is to periodically compare the weights of the first and last tap, and then shift tap weights by one in the direction of the least weight.

9.2 Special-case convergence

Equalization when accurate estimates of the data are not available has been discussed for a very different algorithm in Ref. 5. That paper analyzes the case where the channel is perfect but the equalizer is misaligned. It is shown that, for certain tap initializations, convergence to an undesirable setting is possible. For the Godard algorithm, there is no problem when the channel is perfect and the equalizer is misaligned (arbitrarily). To show that there is no difficulty, we can make use of the analysis in Section V. In this case, $s_j = c_j$. The finiteness of

the number of taps does not alter the argument. Convergence to the optimum tap setting is guaranteed.

For the second example, in the potential function of (4), interpret s (and h and c) as a discrete Fourier transform. Then it is not difficult to show that, within this modular context, convergence of c occurs. This is an artificial construct. However, a variation of this example of a finite tap equalizer may have some utility. Prospective application is not within the scope of equalization procedures of the kind that leave the transmitted signal inviolate. Rather, the application is associated with the equalizer booting methods of the kind that use intervals of periodic training sequences. (The DFT is the key to modeling such equalizers.[16]) The motivation for using quartic criteria such as in eq. (3) is that an immunity to frequency offset is anticipated. This type of equalization, which uses both a quartic potential and training sequence, has arisen in concurrent research by A. A. M. Saleh and the author.

9.3 Status

The issue of the effect of limiting the number of taps is, at bottom, the question of whether with sufficiently many taps and with centering as in Section 9.1 the evolution of the quartic error departs negligibly from the infinite tap case. Godard[1] does not discuss the issue of limiting the number of taps.

For a given application, one could circumscribe a universal ensemble of desired h^{-1} and then choose the number of taps large enough so the approximation error is uniformly negligible, i.e., so that the omitted taps are essentially zero anyway. At each point in time, the taps that can evolve do so in such a way that the infinitesimal tap evolution is the same as the unlimited tap case. The taps that are omitted are essentially preset at their optimum values ($\approx zero$). The random component of the evolution serves to mask the effect of any perturbations of the infinite tap vector field that is caused by limiting the number of taps.

The material presented so far supports that, with a sufficiently large number of taps (and employing a weight centering procedure), the convergence properties of the Godard algorithm differ negligibly from what is predicted for the infinite tap equalizer. However cogent the supporting evidence, a mathematically rigorous argument has not been given. The understanding of the finite tap issue is refined by a numerical example that concludes this section.

9.4 ISI versus number of taps

With the aid of the computer one can take a deeper look at finite tap behavior. A computer program has been written to solve the *deterministic* equation of tap evolution

$$c_{[n+1]} = c_{[n]} - \lambda \nabla \phi(c_{[n]}).$$

One can track the MSE from start, $\text{MSE}_{[0]}$, to equilibrium, $\text{MSE}_{[\infty]}$. The potential function, ϕ, could be MMSE (with known data) or the quartic (with unknown data). The number of levels and the timing epoch are potential function parameters. For numerical work, some of the idealizing assumptions were dropped and we generalized to incorporate important practical effects. Namely, the equalizer can have fractionally spaced taps and the transmitted pulse can be raised cosine with arbitrary roll-off factor.

It is beyond the scope of this paper to include a comprehensive numerical study. We do, however, report the result of one interesting computer experiment. Figure 3(a) expresses $\text{MSE}_{[\infty]}$ versus the number of taps for some anecdotal cases of multipath fade. The fade characteristics are in accordance with the Rummler model[11] for multipath with midband notches in the range 16 to 22 dB. The Rummler phase parameter is zero and the scale parameter is inconsequential, as the affect of additive noise is neglected. See Fig. 3b. The roll-off is 25 percent, and the fractional spacing is half that of synchronous equalization. The timing epoch is optimized. The curves were generated for $L = \infty$ (as we are primarily interested in large constellations and the characteristics are insensitive to changes in L for L large). Computations were made for 3, 5, 7, and 9 taps. The curves shown are solid merely for interpolating to the even ordinates; of course there is no meaning to a noninteger number of taps.

The original intent in generating the characteristics in Fig. 3 was to compare $\text{MSE}_{[\infty]}$ for the quartic and quadratic potentials. However, once the points were determined it was discovered that there was no difference observable to the eye! In this computer experiment the quartic equilibrium is essentially as good as the quadratic one in minimizing MSE. This virtual equality is stronger than what is required of the quartic algorithm. We only desire that the quartic equilibrium be close enough to the quadratic equilibrium so as to create the option of switching to MMSE evolution once good decisions are available.

X. DISCUSSION

We have delved into the theoretical aspects of quartic algorithm. The results obtained included stability with respect to tap initialization (Section V), equalization including acquisition of phase in systems employing highly stable oscillators (Section VII), and cancellation of crosspolarization interference (Section VIII). Section VI discussed convergence times which are of interest in bootstrapping digital radio systems. Section IX suggests that with centering, and with enough

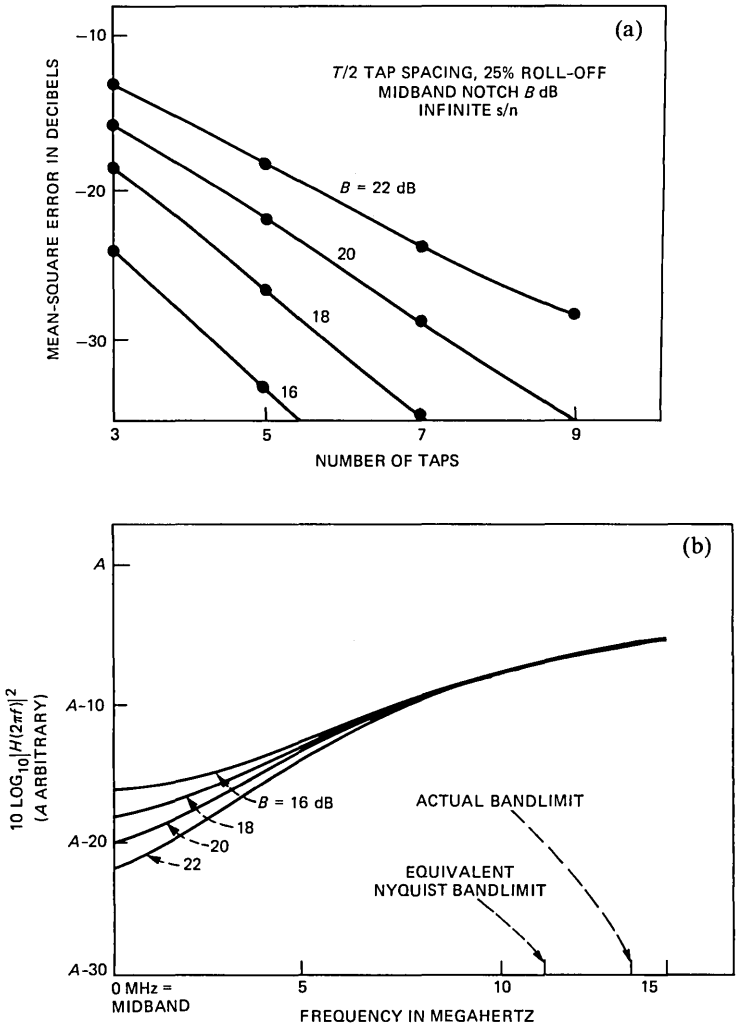


Fig. 3—(a) Steady-state mean-square error versus number of taps for quartic criterion. (b) Power transfer characteristics of channels used in computation of mean-square $error_{[\infty]}$ for quartic algorithm.

taps, the finite tap equalizer may perform essentially as well as the infinite tap equalizer.

It is premature to answer the question of whether digital radio systems should be designed to provide for self-alignment without inserting media probing signals in the transmitted signal. However, the results that have been elucidated thus far are promising enough to warrant further study.

The electronics needed for implementation needs to be assessed.

Ideally, one would like to update taps every symbol interval. Then convergence speed is high and the operation of the algorithm is less vulnerable to the assumption that the channel is unchanged during booting. It is conceivable to implement such an algorithm with special-purpose hardware. However, this now may prove unrealistic from an economic standpoint. (In time the economic issue will disappear.) One could slow the algorithm, updating every 10 or 100 symbols, to obtain a more realistic implementation. By slowing the algorithm, convergence speed, rather than economics, becomes the issue. The question of how fast an algorithm is needed is particularly difficult to address, especially for cross-polarization cancellation, because of the lack of data on coupling dynamics. One could strive to compensate for slowness by developing good, adaptive, step-size algorithms. These are all items for future study.

The best approach for future investigations of the usefulness of quartic algorithms in digital radio would seem to require inclusion of simulation and/or experimentation aimed at specific applications. Ultimately, it is necessary to include additive noise in such evaluations.

XI. ACKNOWLEDGMENTS

The author gratefully acknowledges helpful interactions with N. Amitay, R. D. Gitlin, L. J. Greenstein, M. I. Reiman, A. A. M. Saleh, J. Salz, A. Weiss, J. J. Werner, and Y. S. Yeh.

REFERENCES

1. D. Godard, "Self-Recovering Equalization and Carrier Tracking in Two-Dimensional Data Communication Systems," IEEE Trans. Commun., *COM-28* (November 1980), pp. 1867–75.
2. D. Godard, "A 9600 Bit/s Modem for Multipoint Communications Systems," IEEE NTC 1981 Conf. Rec., New Orleans, LA., November 29–December 3, pp. B3.3.1–B3.3.5.
3. Y. Sato, "A Method of Self-Recovering Equalization for Multilevel Amplitude-Modulation Systems," IEEE Trans. Commun., *COM-23* (June 1975), pp. 679–82.
4. J. E. Mazo, "Analysis of Decision-Directed Equalizer Convergence," B.S.T.J., *59,* No 10 (December 1980), pp. 1857–76.
5. A. Beneniste and M. Goursat, "A Gain and Phase Identification Procedure: Blind Adjustment of a Recursive Equalizer," IEEE Inform. Theory Symp., Grignano, Italy, June 1980.
6. A. Benveniste and M. Goursat, "Blind Equalizers," INRIA, Rapports de Richerche, No. 219, July 1983.
7. G. J. Foschini and J. Salz, "Digital Communications Over Fading Radio Channels," B.S.T.J., *62,* No. 2, Part 1 (February 1983), pp. 429–56.
8. J. Salz, "Optimum Mean-Square Decision Feedback Equalization," B.S.T.J., *52,* No. 8 (October 1973), pp. 1341–73.
9. D. D. Falconer and G. J. Foschini, "Theory of Minimum Mean-Square-Error QAM Systems Employing Decision Feedback Equalization," B.S.T.J., *52,* No. 10 (December 1973), pp. 1821–48.
10. A. C. Salazar, "Design of Transmitter and Receiver Filters for Decision Feedback Equalization," B.S.T.J., *53,* No. 3 (March 1974), pp. 503–23.
11. W. D. Rummler, "A New Selective Fading Model: Application to Propagation Data," B.S.T.J., *58,* No. 5 (May–June 1979), pp. 1032–71.
12. B. Widrow and J. M. McCool, "A Comparison of Adaptive Algorithms Based on the

Methods of Steepest Descent and Random Search," IEEE Trans. Antennas and Propagation, *AP-24*, No. 5 (September 1976), pp. 615–37.
13. S. Karlin and H. M. Taylor, *A Second Course in Stochastic Process*, New York: Academic Press, 1981, Chapter 15.
14. L. Arnold, *Stochastic Differential Equations*, New York: Wiley-Interscience, 1974.
15. N. Amitay and L. J. Greenstein, "Multipath Outage Performance of Digital Radio Receivers Using Finite-Tap Adaptive Equalizers," IEEE Trans. Commun., *COM-32* (May 1984), pp. 597–608.
16. K. H. Mueller and D. A. Spaulding, "Cyclic Equalization—A New Rapidly Converging Equalization Technique for Synchronous Data Communication," B.S.T.J., *54*, No. 2 (February 1975), pp. 370–406.

APPENDIX

Even-Order Moments of QAM Constellations

The moments $E|a_i|^4$ for constellations with L^2 points are required in the paper. A general procedure for calculating $E|a_i|^N$ (N even) is as follows: Write a recursion in L for $m_N(L) \triangleq E(\text{Re } a_i)^N$. The recursion can be solved using transforms. Straightforward algebra can be used to obtain $E|a_i|^N$ from $m_N(L)$, e.g.,

$$E|a_i|^2 = 2m_2$$

$$E|a_i|^4 = (m_4 + m_2^2)$$

$$E|a_i|^6 = (m_6 + 3m_4m_2)$$

$$E|a_i|^8 = 2(m_8 + 4m_6m_2 + 3m_4^2).$$

Normalizing $E|a_i|^2 = 1$ and following the above suggestion for $N = 4$ gives

$$E|a_i|^4 = \frac{7L^2 - 13}{5(L^2 - 1)}.$$

For a constellation with a large number of points it is useful to have the asymptotic ($L \to \infty$) form for the moments. The result is

$$\lim_{L\to\infty} m_N(L) = \frac{1}{2A}\int_{-A}^{A} |\zeta|^N d\zeta = \frac{A^N}{N+1}.$$

Since $E|a_i|^2 = 1$ we have $A = \sqrt{3/2}$. Therefore

$$E|a_i|^4 \to 7/5$$

$$E|a_i|^6 \to 81/35$$

and $$E|a_i|^8 \to 747/175.$$

These asymptotes were used in composing (7b). As a check compare $E|a_i|^4$ in the exact and asymptotic form. Note $E|a_i|^4$ asymptotes quickly. For $L^2 = 16$ points $E|a_i|^4 = 1.32$ as compared to 1.4 for an infinite point constellation.

AUTHOR

Gerard J. Foschini, B.S.E.E., 1961, Newark College of Engineering, Newark, NJ; M.E.E., 1963, New York University, New York; Ph.D. (Mathematics), 1967, Stevens Institute of Technology, Hoboken, NJ; AT&T Bell Laboratories, 1961—. Mr. Foschini initially worked on real-time program design. For many years he worked in the area of communication theory. In the spring of 1979 he taught at Princeton University. Mr. Foschini has supervised planning the architecture of data communications networks. Currently, he is involved with optical communication research. Member, Sigma Xi, Mathematical Association of America, IEEE, New York Academy of Sciences.

AT&T Technical Journal
Vol. 64, No. 8, October 1985
Printed in U.S.A.

Baseband Cross-Polarization Interference Cancellation for M-Quadrature Amplitude-Modulated Signals Over Multipath Fading Radio Channels

By M. KAVEHRAD*

(Manuscript received March 21, 1985)

In this paper we propose a novel baseband structure capable of adaptively mitigating cross-polarization interference in a dual-polarized, M-state quadrature amplitude-modulated received signal. We show that by using this canceler, performance signatures very close to single-polarized system signatures can be achieved for dually polarized digital radio systems.

I. INTRODUCTION

Because of frequency reuse via orthogonally polarized channels, dual-polarized transmission of M-state Quadrature Amplitude-Modulated (M-QAM) signals can double the bandwidth efficiency of terrestrial radio routes. Such systems transmit two different information signals of the same bandwidth and the same carrier frequency by using orthogonal field polarization for the transmission of each signal. Nonideal antennas and transmission media cause cross-coupling of the two signals and cross-polarization interference. Cross-polarization interference cancellation using adaptive transversal filters over linear dispersive multipath channels has been the subject of considerable prior investigation.[1-4]

In this study we deal with cross-polarization interference cancella-

* AT&T Bell Laboratories.

tion and intersymbol interferance (ISI) equalization separately, and propose a novel method of cross-polarization cancellation for dual-polarized operation of M-QAM signals over dispersive fading channels, similar to those experienced in line-of-sight terrestrial radio applications. The canceler operates at baseband and improves the dual-polarized system performance to very nearly the performance of a single-polarized system.

The canceler design is based on a previous observation that the power loss associated with a cross-coupled signal subject to flat or mildly dispersive fading brings about an actual reduction in system outage time.[5] In this paper we use the model and results of Ref. 5 to introduce the canceler structure and evaluate its performance. To enable comparison in the absence of cancellation, we use a dual-polarized system performance signature (M-curve) as a measure in our evaluation.[5]

In the following section we review the results of Ref. 5 briefly, and then introduce discussions leading to the realization of the canceler. In Section III the canceler performance for both dual-polarized 16- and 64-QAM radio systems is presented and the results are discussed in detail.

II. ANALYTICAL MODEL

In this section we describe briefly the channel model and underlying assumptions germane to baseband cancellation, and then introduce the canceler model.

2.1 Channel model

The dual-polarized channel model is shown in Fig. 1. Two inde-

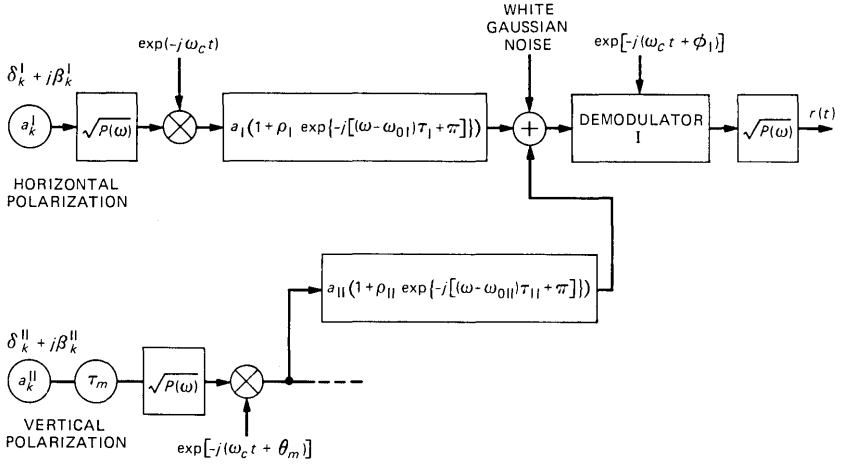


Fig. 1—Channel model for a dual-polarized system.

pendent data streams are separately modulated and transmitted co-channel, using orthogonal polarizations. Rummler's multipath fading model, which assumes the presence of a single inband notch, is applied to both the copolarized and the cross-coupled interfering channel transfer functions.[6] In other words, fadings of both the desired copolarized and cross-polarized interference channels are assumed to be the single notch type and independent. As depicted in Fig. 1 and derived in detail in Ref. 5, the received in-phase signal on the reference copolarized path is denoted by $r_{i,\mathrm{I}}(t)$;

$$
\begin{aligned}
r_{i,\mathrm{I}}(t) = {} & a_{\mathrm{I}}\delta_0^{\mathrm{I}}\{\cos(\phi_{\mathrm{I}})p(t) + \rho_{\mathrm{I}}\cos[(\omega_c - \omega_{0\mathrm{I}})\tau_{\mathrm{I}} + \pi + \phi_{\mathrm{I}}]p(t - \tau_{\mathrm{I}})\} \\
& + a_{\mathrm{I}} \sum_{k\neq 0} \delta_k^{\mathrm{I}}\{\cos(\phi_{\mathrm{I}})p(t - kT_s) + \rho_{\mathrm{I}}\cos[(\omega_c - \omega_{0\mathrm{I}})\tau_{\mathrm{I}} + \pi + \phi_{\mathrm{I}}] \\
& \quad \cdot p(t - kT_s - \tau_{\mathrm{I}})\} \\
& + a_{\mathrm{I}} \sum_{k} \beta_k^{\mathrm{I}}\{\sin(\phi_{\mathrm{I}})p(t - kT_s) + \rho_{\mathrm{I}}\sin[(\omega_c - \omega_{0\mathrm{I}})\tau_{\mathrm{I}} + \pi + \phi_{\mathrm{I}}] \\
& \quad \cdot p(t - kT_s - \tau_{\mathrm{I}})\} \\
& + a_{\mathrm{II}} \sum_{k} \delta_k^{\mathrm{II}}\{\cos(\phi_{\mathrm{I}} + \theta_m)p(t - kT_s - \tau_m) \\
& + \rho_{\mathrm{II}}\cos[(\omega_c - \omega_{0\mathrm{II}})\tau_{\mathrm{II}} + \pi + \phi_{\mathrm{I}} + \theta_m]p(t - kT_s - \tau_{\mathrm{II}} - \tau_m)\} \\
& + a_{\mathrm{II}} \sum_{k} \beta_k^{\mathrm{II}}\{\sin(\phi_{\mathrm{I}} + \theta_m)p(t - kT_s - \tau_m) \\
& + \rho_{\mathrm{II}}\sin[(\omega_c - \omega_{0\mathrm{II}})\tau_{\mathrm{II}} + \pi + \phi_{\mathrm{I}} + \theta_m]p(t - kT_s - \tau_{\mathrm{II}} - \tau_m)\} \\
& + \mathrm{Re}\{n_{\mathrm{I}}(t)\}, \qquad (1)
\end{aligned}
$$

where $\mathrm{Re}\{\cdot\}$ stands for real part. In this equation, (δ_k^i, β_k^i) $i = \mathrm{I}, \mathrm{II}$ represent the real and imaginary parts of the complex-valued transmitted symbols on the two polarizations, I and II, at consecutive instants, kT_s, $k = 0, 1, 2, \cdots$, where T_s is a baud period. The Nyquist-shaping filter impulse response is denoted by $p(t)$, and ω_c is the nominal carrier frequency. The parameters a_i, ρ_i, ω_{0i}, τ_i; $i = \mathrm{I}, \mathrm{II}$ represent the flat fade level, fade notch depth, fade notch position, and relative delay between the two rays in each of the Rummler type multipath fading models (the reference copolarized and the corresponding cross-coupled interfering channels). Also, in eq. (1), τ_m and θ_m account for any symbol timing or carrier phase asynchronism that may exist between the two polarized signals at the transmitter location.

It might be worthwhile, at this point, to discuss the impact of transmitter local oscillators status on the theoretical modeling of the channel. Illustrated in Fig. 2 is a typical dual-polarized system transmitter configuration. As seen, there are three major sets of local oscillators in the transmitter system that can play an important role

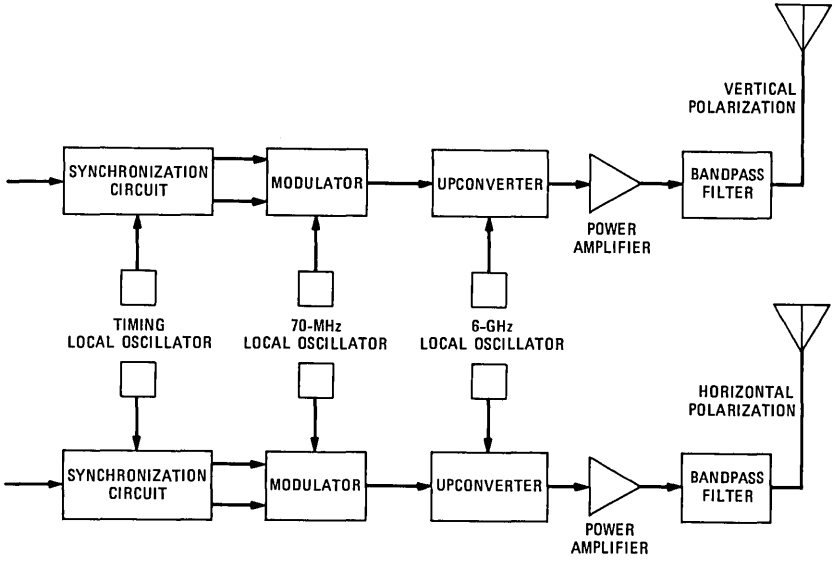


Fig. 2—A dual-polarized system transmitter.

in modeling a dual-polarized system. Namely, local oscillators used to provide clock timing to baseband sequences, IF local oscillators providing carrier signals to modulators, and microwave upconverter oscillators. Because we intend to introduce baseband cancellation in the following sections, receiver implementation is simpler if we synchronize all the transmitter local oscillators. In other words, we assume $\tau_m = 0$ and $\theta_m = 0$ and investigate the overall system performance. This assumption also results in improved performance of the cross-polarization interference canceler as was demonstrated in Ref. 5 for the general dual-polarized system performance signatures in the absence of a canceler.

The optimum phase between the modulator and the demodulator of the reference copolarized signal ($i = \mathrm{I}$) for optimum timing is introduced by ϕ_{I}. Note that for a strong main polarization signal and because of the independence assumption on the cross-coupled signal, ϕ_{I} is imposed on the latter by the copolarized signal demodulator.[5]

As noted in eq. (1), the dispersive nature of the multipath channel is completely described by the superposition of four impulse responses, each weighted by an appropriate independent transmitted symbol state. These impulse responses for the kth transmitted symbol intervals are

$$u_{i,\mathrm{I}} = a_{\mathrm{I}}\{p(t - kT_s)\cos(\phi_{\mathrm{I}}) + \rho_{\mathrm{I}}p(t - kT_s - \tau_{\mathrm{I}})\cos[(\omega_c - \omega_{0\mathrm{I}})\tau_{\mathrm{I}} + \pi + \phi_{\mathrm{I}}]\}, \quad (2a)$$

$$u_{q,\mathrm{I}} = a_{\mathrm{I}}\{p(t - kT_s)\sin(\phi_{\mathrm{I}}) + \rho_{\mathrm{I}}p(t - kT_s - \tau_{\mathrm{I}})\sin[(\omega_c - \omega_{0\mathrm{I}})\tau_{\mathrm{I}} + \pi + \phi_{\mathrm{I}}]\}, \quad (2b)$$

$$u_{i,\mathrm{II}} = a_{\mathrm{II}}\{p(t - kT_s - \tau_m)\cos(\phi_{\mathrm{I}} + \theta_m) + \rho_{\mathrm{II}}p(t - kT_s - \tau_{\mathrm{II}} - \tau_m) \cdot\cos[(\omega_c - \omega_{0\mathrm{II}})\tau_{\mathrm{II}} + \pi + \phi_{\mathrm{I}} + \theta_m]\}, \quad (2c)$$

and

$$u_{q,\mathrm{II}} = a_{\mathrm{II}}\{p(t - kT_s - \tau_m)\sin(\phi_{\mathrm{I}} + \theta_m) + \rho_{\mathrm{II}}p(t - kT_s - \tau_{\mathrm{II}} - \tau_m)\sin[(\omega_c - \omega_{0\mathrm{II}})\tau_{\mathrm{II}} + \pi + \phi_{\mathrm{I}} + \theta_m]\}, \quad (2d)$$

where the variables have been previously defined. For the received in-phase part of the main polarization signal, eqs. (2a) and (2b) describe the distorted in-phase and quadrature-coupled signals of the reference copolarized transmitter, respectively, and equations (2c) and (2d) describe the corresponding signals from the cross-polarized interferer.

To introduce the parameters that define the fading character of the interfering cross-coupled signal path, we associated with each interferer fading event a triplet representing its dispersive fading status. This triplet is

$$\left[20 \log \frac{a_{\mathrm{II}}}{a_{\mathrm{I}}}(\mathrm{dB}), -20 \log |1 - \rho_{\mathrm{II}}|(\mathrm{dB}), \Delta f_{0\mathrm{II}}(\mathrm{MHz})\right], \quad (3)$$

where a_{II} and a_{I} represent the flat fade levels for cross-coupled and copolarized signals, respectively. In the triplet, the second term is dispersive fade notch depth, and $\Delta f_{0\mathrm{II}}$ denotes fade notch position relative to the carrier frequency of the cross-coupled channel. (Notice that other definitions of notch depth can be found in the literature.[4]) For illustrative purposes, we demonstrate eqs. (2a) through (2d) in Figs. 3 and 4, an interferer of (−20, 0, 0) fade and two different fade conditions of the reference main polarization path (copolarized channel). In Fig. 3 we illustrate the aforementioned impulse responses when a notch-centered fade of 10 dB is applied to the main polarization signal. Observe that since the fade on the latter is notch centered and $\theta_m = 0$, $u_{q,\mathrm{I}}$ and $u_{q,\mathrm{II}}$ are both zero. In Fig. 4 an 11-MHz offset fade of 7.5 dB is applied to the main polarization path, and even though the interferer has a flat fade, because of the phase ϕ_{I} imposed on it, $u_{i,\mathrm{II}}$ and $u_{q,\mathrm{II}}$ are nonzero Nyquist-shaped pulses with their relative positions also determined by the phase and timing imposed on them by the dominant polarization signal.

Now we define a decision variable which is a function of the desired symbol to be detected, intersymbol interference, cross-polarized interference, and Gaussian noise. To evaluate the average error probability, first we derive the conditional error probability conditioned on the

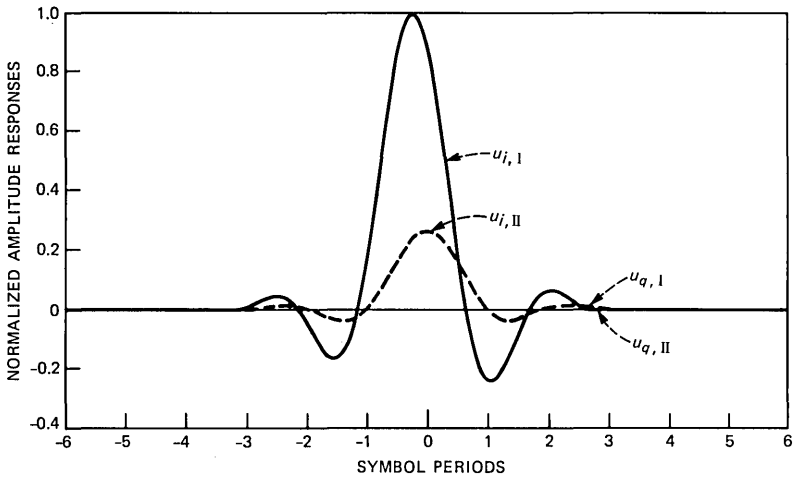


Fig. 3—16-QAM signal, time-domain impulse responses for a notch-centered fade.

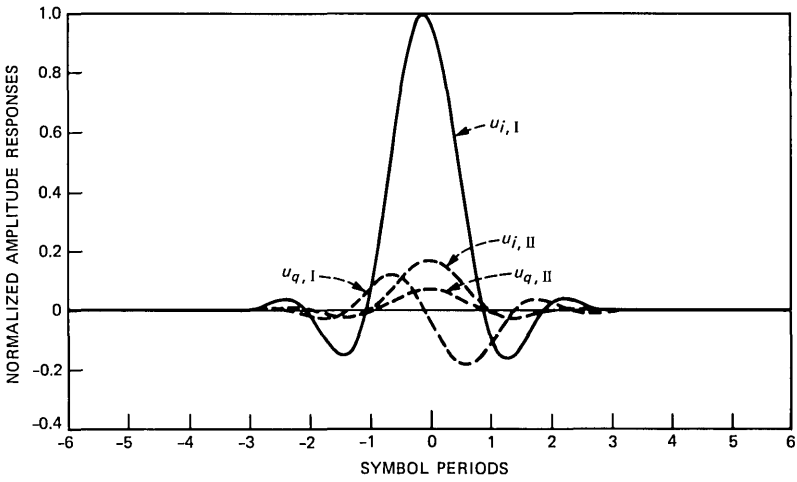


Fig. 4—16-QAM signal, time-domain impulse responses for an offset fade.

composite interference. Then by applying moment-generating functions and the Gauss quadrature method, we determine the average error probability. The details of this procedure are explained in Ref. 5.

Using eq. (1), we then computed the performance signatures (M-curves) of the main (reference) polarization signal (i = I) that provide a locus of the fade notch depth (in dB) versus the relative fade notch position (in MHz) for a 10^{-3} average probability of error. In Fig. 5 we illustrate the performance signatures of the main polarization

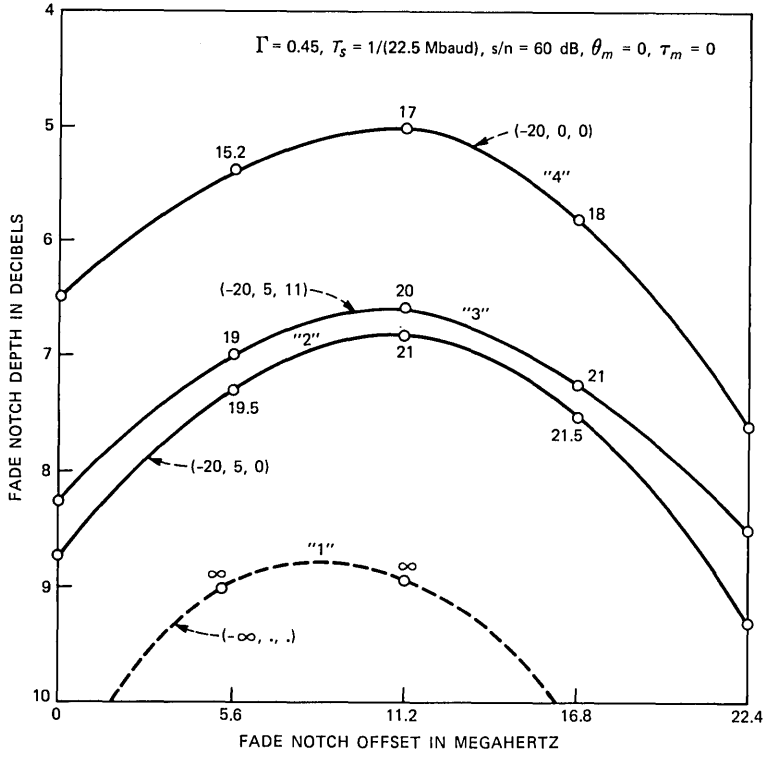


Fig. 5—Performance signature curves for dual-polarized 16-QAM radio.

signal, that is, $-20 \log |1 - \rho_I|$ versus Δf_{0I}, where ρ_I is the dispersive fade notch depth of the main polarization path, and Δf_{0I} denotes its fade notch position relative to the carrier frequency. Along the curves we have specified average signal-to-interference ratio at a selected number of points. As a reference we illustrate the signature of a single-polarized 16-QAM system, that is, $a_{II} = 0$ and label it "1." A comparison of curves labeled "2" through "4" for different fadings of the interferer in Fig. 5 reveals the aforementioned fact that the system outage time (area under the M-curve) is related to the net interfering power for a mild dispersive fading of the interfering signal on the cross-coupled path. For example, a comparison of curves 4 and 2 with the same 20-dB flat power levels and 0-MHz notch offsets reveals that curve 2 with a 5-dB inband notch fade, results in *less* outage time than the fade of curve 4 with no inband notch. Hence, the greater power loss associated with curve 2 leads to reduced outage, even though the intersymbol interference at the reference receiver in the case of curve 2 exceeds that of curve 4.

Now consider curves 2 and 3. The data corresponds to identical flat

power levels and fade notch depths, with the notch position moving from 0 MHz (notch centered) to 11 MHz (near the band edge). The notch-centered fade causes less outage than the notch offset fade because the unfaded signal spectral energy at 0 MHz is much more than that near the band edge; hence, the relationship of curves 3 and 2 is again that of diminished net signal power in the interferer resulting in a reduced outage. All these curves were computed for a 60-dB signal-to-noise ratio (s/n), 22.5-Mbaud symbol rate, $\Gamma = 0.45$ Nyquist filter roll-off, and a 16-QAM radio system.

2.2 Canceler model

It is well known that interference power is directly related to the area of the cross-coupled signal power spectral density. Thus, in dual-polarized operation, where the dual-polarized signals are transmitted cochannel, any reduction of the interfering signal power spectral density area leads to a decrease in the overlap area between the main and the cross-coupled signal spectral densities, and, hence, a reduction in the interfering power. Therefore, a cross-polarized interference canceler able to perform such a task should bring about an improvement in the performance of the dual-polarized system. It is also well known that the main lobe sample of a Nyquist-type pulse is proportional to the area of its frequency spectrum. Owing to this fact, we hypothesized improvements in the dual-polarized system-performance signatures, given that the main lobe of the cross-coupled interferer is canceled in time domain. This hypothesis proved to be correct and is discussed further in Section III.

A block diagram of the cross-polarized interference canceler and system equalizers is shown in Fig. 6. Decision feedback complex taps cancel the main lobe of the cross-coupled interfering signal adaptively, using preliminary estimates of the main-lobe. Least-Mean-Square (LMS) adaptation is recommended because it was shown[7] that s/n degradation by some cross-polarization cancellation methods can be large and that the adaptive algorithm should take into account noise power minimization.[7] This is known to be one of the salient features of the LMS algorithms.[8] Because the proposed cancellation is performed at baseband, the difference between input and output of the detector slicer circuit (error signal) can be employed as the performance measure and it can be utilized by the LMS controller to derive the canceler coefficients.

Note that in Fig. 6 the baseband canceler precedes the system equalizers. This is to prevent the equalizers from causing excessive dispersion in the interfering signal when attempting to equalize deep fade notches of the copolarized signal as is the case in combined cross-polarized cancellation and ISI equalization. Note that the system

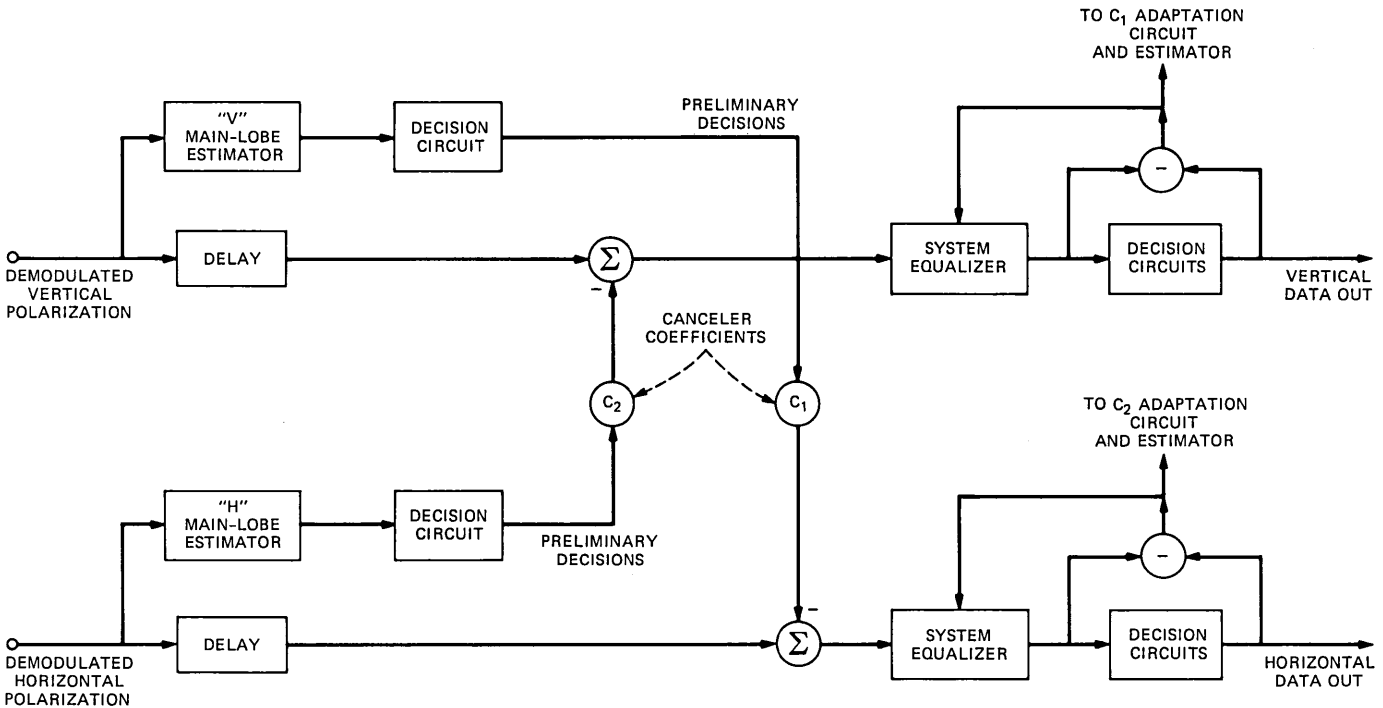


Fig. 6—Cross-polarized canceler/equalizer structure.

equalizers are used to mitigate intersymbol interference and cross-rail interference; therefore, they are not parts of the cross-polarization cancellation operation.

To perform main-lobe cancellation in practice, preliminary estimates of the signal main lobe on one polarization must be subtracted from the opposite polarization received signal that needs to be delayed by the amount of time that it takes to estimate the lobe.

In this theoretical study we assume the preliminary estimates of the signal main lobe are correct so that we can cancel the cross-polarized signal. In practice this assumption is valid in the steady-state mode of operation if some kind of bootstrapped algorithm is adopted. This, of course, adds to the circuit complexity. An obvious advantage of this method is that the preliminary decisions used to cancel the cross-polarized signal provide noise-free estimates; hence, less s/n degradation is caused by coupling mechanism compared to feedforward methods.

In the next section we elaborate on the system performance signatures after cross-polarization interference cancellation as well as making comparisons to the signatures of the same system without cross-polarization cancellation that we use as base-line measures.

III. CANCELER PERFORMANCE

In this section we present the computed performance signature curves for dual-polarized M-QAM signals using the canceler described in the previous section.

Results in the form of performance signatures are illustrated in Figs. 7 through 10. As we can observe, use of a single complex decision feedback tap to cancel the real and imaginary parts of the cross-coupled interferer main-lobe sample renders performance signatures practically identical to those of a single-polarized system, in dual-polarized operation. To elaborate on the required number of canceler taps, it should be obvious in this case that only a single complex tap is adequate to remove the main lobe of the interferer. This is because when there is no fading or when there is offset fading of the copolarized channel, the interferer main lobe always coincides, or approximately coincides, with the desired symbol main lobe, and only one complex feedback tap is necessary to remove it. However, in the case of midband fading of the copolarized signal, as seen in Fig. 7, the timing reference of the main path impulse response is offset from the peak of the main lobe of the interfering signal. Hence, the canceler does not perform as well for midband fades as it would for offset fades of the copolarized path. This is because, for offset fades, as the copolarized path fade notch moves toward the band edge, the timing reference of the overall impulse response moves toward the origin;[5] hence, the two peaks tend

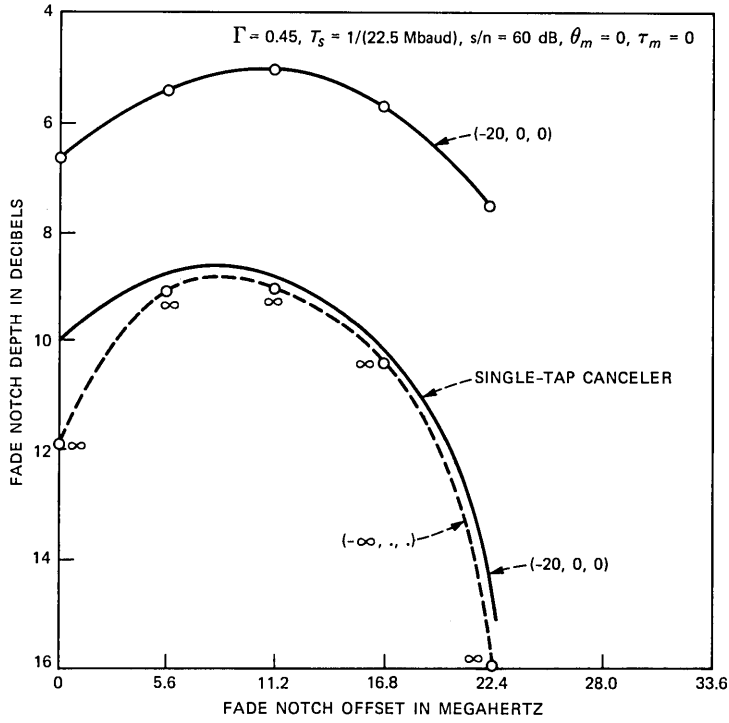


Fig. 7—Canceler performance in dual-polarized 16-QAM radio for a flat fade on cross-coupled interfering path.

to align. Of course, the remedy for the centered fade situation is to increase the number of the canceler transversal taps.

In Figs. 7 and 8, all the curves were computed for a 60-dB s/n, 22.5-Mbaud symbol rate, $\Gamma = 0.45$ roll-off, 16-QAM radio; and in Figs. 9 and 10, the signatures for 64-QAM radio were computed for a 66-dB s/n, 15-Mbaud symbol rate, and $\Gamma = 0.45$ roll-off. *Note that, in all these computations, ISI equalization of the main polarization signal is left out.*

To further quantify the influence of interfering signal main lobe on the dual-polarized system outage performance, we present an example. In the 16-QAM radio case, for an interfering signal defined by the triplet (−20, 5, 0) and for a centered fade of 6-dB notch depth on the reference copolarized signal, samples taken from $u_{i,\mathrm{I}}$, $u_{q,\mathrm{I}}$, $u_{i,\mathrm{II}}$, and $u_{q,\mathrm{II}}$ at optimum timing points are listed in Table I. As observed, the peak sample of the interferer impulse responses, $u_{i,\mathrm{II}}$ and $u_{q,\mathrm{II}}$, have amplitudes about ten times larger than the sum of absolute values of all their other samples taken every baud period. Note that, since θ_m and τ_m are zero, the imaginary part of the interferer impulse response, $u_{q,\mathrm{II}}$,

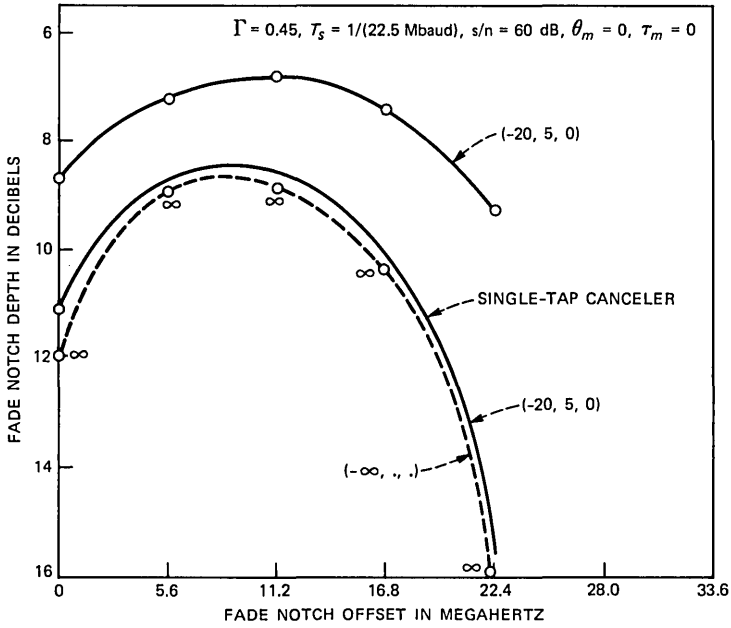


Fig. 8—Canceler performance in dual-polarized 16-QAM radio for a dispersive fade on cross-coupled interfering path.

is zero because of the notch-centered fading of the main polarization signal in this case, that is, $\phi_1 = 0$ (see Fig. 3). To ensure validity of the test, for the same interferer fading conditions, we repeated this for several different fading conditions of the copolarized signal and checked the resulting impulse responses, $u_{i,\mathrm{II}}$ and $u_{q,\mathrm{II}}$. In all the cases considered, samples of the real and imaginary parts of the interfering signal main lobe ranged somewhere between eight to ten times the value of the sum of the absolute values of all other samples. Thus, the contribution of the main-lobe sample of the interferer to the total peak distortion is much stronger than that of all other samples. Therefore, canceling the real and imaginary parts of the interferer main lobe should improve the performance significantly, as expected.

IV. CONCLUSIONS

In this paper we proposed a novel baseband cross-polarization interference canceler structure that adaptively mitigates the interference in a dual-polarized M-QAM radio system. Employing performance signatures (M-curves) for dual-polarized systems, as introduced in Ref. 5, we showed that for synchronous transmitters, a single-decision feedback complex matrix tap canceler can enhance a dual-polarized system availability time to values close to the availability

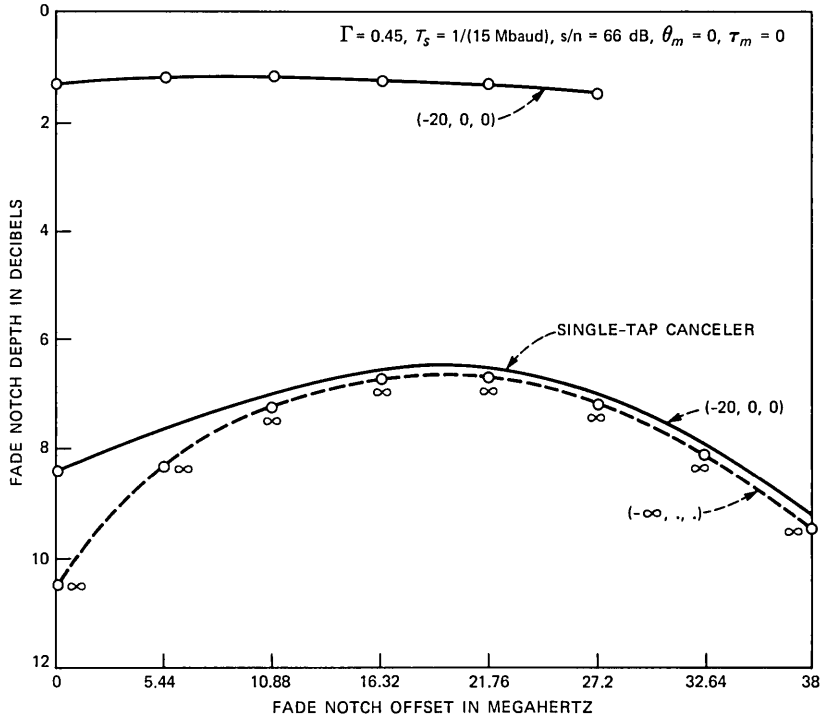


Fig. 9—Canceler performance in dual-polarized 64-QAM radio for a 20-dB flat fade on cross-coupled interfering path.

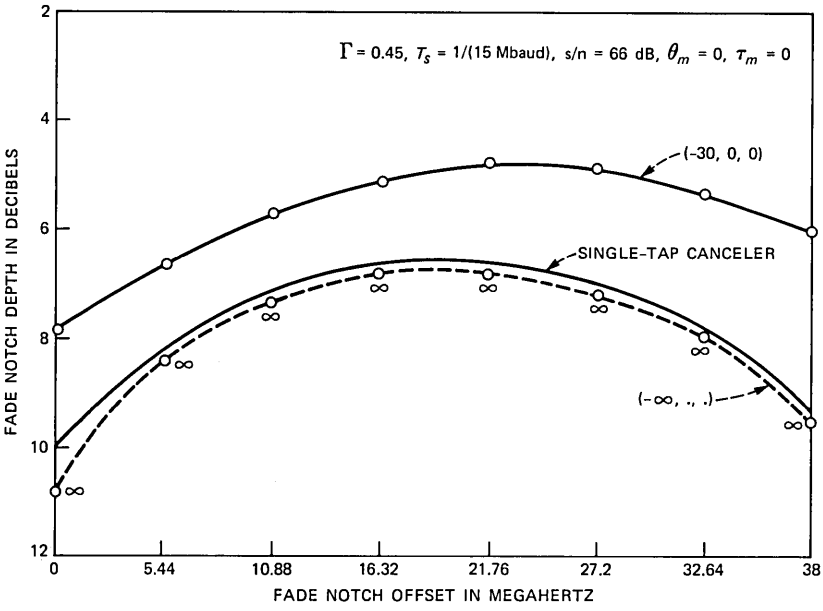


Fig. 10—Canceler performance in dual-polarized 64-QAM radio for a 30-dB flat fade on cross-coupled interfering path.

Table I—Impulse response samples of received in-phase signal on reference polarization

$u_{i,I}$	$u_{q,I}$	$u_{i,II}$	$u_{q,II}$	Time
0.5462E-04	0.3645E-19	0.7532E-05	0.1074E-19	$t_0 + 9T_s$
0.1560E-04	0.1690E-18	−0.7399E-06	0.9008E-20	.
−0.1268E-03	−0.1694E-18	−0.1594E-04	−0.2812E-19	.
0.3483E-04	−0.2848E-18	0.1042E-04	−0.4691E-20	.
0.3412E-03	0.6616E-18	0.3914E-04	0.8339E-19	.
−0.3989E-03	0.3152E-18	−0.6560E-04	−0.5668E-19	.
−0.1374E-02	−0.3568E-17	−0.1411E-03	−0.3695E-18	.
0.7526E-02	0.9137E-17	0.9628E-03	0.1635E-17	$t_0 + 2T_s$
−0.2679E-01	−0.2553E-16	−0.3555E-02	−0.5556E-17	$t_0 + T_s$
0.5328E+00	0.1171E-14	0.5860E-01	0.1354E-15	t_0
−0.1332E-01	−0.1469E-15	0.6797E-03	−0.7793E-17	$t_0 - T_s$
0.2236E-02	0.3376E-16	−0.2800E-03	0.1649E-17	$t_0 - 2T_s$
−0.1519E-02	−0.9138E-17	−0.6131E-04	−0.6031E-18	.
0.5787E-03	0.4275E-18	0.7905E-04	0.1154E-18	.
0.8068E-04	0.1226E-17	−0.1024E-04	−0.5979E-19	.
−0.1865E-03	−0.3837E-18	−0.2098E-04	−0.4639E-19	.
0.2264E-04	−0.2996E-18	0.8858E-05	−0.7337E-20	.
0.7182E-04	0.2201E-18	0.6764E-05	0.2058E-19	.
−0.2951E-04	0.7752E-19	−0.5842E-05	−0.2163E-20	$t_0 - 9T_s$

time for a single-polarized radio system, for the assumed propagation model.

REFERENCES

1. C. A. Baird and G. Pelchat, "Cross Polarization Techniques Investigation," Harris Corporation Report No. RADC-TR-77-244, July 1977.
2. B. E. Gillingham et al., "Cross Polarization Interference Reduction Techniques," Harris Corporation Report No. RADC-TR-79-154, June 1979.
3. J. Namiki and S. Takahara, "Adaptive Receiver for Cross-Polarized Digital Transmission," Int. Conf. Commun., June 14–18, 1981, Denver, Colorado, Paper 46.3.1.
4. M. L. Steinberger, "Design of a Terrestrial Cross-Pol Canceler," Int. Conf. Commun., June 1982, Philadelphia, pp. 2B.6.1–5.
5. M. Kavehrad and C. A. Siller, private communication.
6. W. D. Rummler, "A New Selective Fading Model: Application to Propagation Data," B.S.T.J., *58*, No. 5 (May–June 1979), pp. 1037–71.
7. M. Kavehrad, "Performance of Cross-Polarized M-ary QAM Signals Over Nondispersive Fading Channels," AT&T Bell Lab. Tech. J., *63* (March 1984), pp. 499–521.
8. J. G. Proakis, *Digital Communications*, New York: McGraw-Hill, 1983.

AUTHOR

Mohsen Kavehrad, B.S. (Electrical Engineering), 1973, Tehran Polytechnic Institute; M.S. (Electrical Engineering), 1975, Worcester Polytechnic Institute; Ph.D. (Electrical Engineering), 1977, Polytechnic Institute of New York; Fairchild Industries, 1977–1978; GTE, 1978–1981; on the faculty of Northeastern University, 1981–1984; AT&T Bell Laboratories, 1981—. At AT&T Bell Laboratories Mr. Kavehrad is a member of the Communications Methods Research Department. His research interests are digital communications and computer networks. He is a Technical Editor for the IEEE Communications Magazine. He established and was the Chairman of the IEEE Communications Chapter of New Hampshire in 1984. Member, IEEE, Sigma Xi.

AT&T Technical Journal
Vol. 64, No. 8, October 1985
Printed in U.S.A.

Performance of Low-Complexity Channel Coding and Diversity for Spread Spectrum in Indoor, Wireless Communication

By M. KAVEHRAD* and P. J. McLANE†

(Manuscript received January 30, 1985)

The application of selection diversity in conjunction with simple channel coding is considered for a multiuser, slowly fading, Spread-Spectrum Multiple Access (SSMA), digital radio system. For the most part, the index of performance for our study is the average bit error probability; we also give some consideration to multipath outage as a performance measure. All subscribers are assumed to communicate to a central station; that is, a star network architecture is assumed. *Average* power control is also assumed. The average mentioned in this context includes averaging over the channel fading statistics. The modulation is direct-sequence, spread-spectrum, binary phase-shift keying. We assume perfect timing and carrier recovery in our coherent receiver, and a slowly varying, Rayleigh fading, discrete multipath model is used. Previous analyses have found that SSMA can tolerate few simultaneous users for fading radio channels. We find that the combination of spread-spectrum modulation with low-complexity diversity and/or channel coding can restore fading-channel user levels to an acceptable figure. In addition, selection diversity plus channel coding is more effective than either method by itself. Finally, it turns out that SSMA is less sensitive to a change in the value of delay spread of a fading channel than, say, time-division multiple access. The method of moments is used to accurately assess the system error probability. Using this technique, we also assess the accuracy of assuming that the multiuser interference has a Gaussian distribution, which allows it to be analyzed by a simple method. Using this assumption, we compare selection

* AT&T Bell Laboratories. † Queen's University, Kingston, Canada.

diversity plus channel coding with the maximal-ratio-combining technique for diversity reception. Except for a high order of diversity, the former is more efficient and is always less complex than the latter.

I. INTRODUCTION

In a recent paper Kavehrad has presented a technique to evaluate the performance of direct-sequence, spread-spectrum binary phase-shift-keying modulation for an Indoor, Wireless Communication (IWC) channel.[1] The analysis uses the method of moments,[2] which gives accurate estimates of error probability for many digital communication systems. Kavehrad did not consider diversity in his study. We find that Kavehrad's formulas are only slightly modified when selection diversity (see pages 313 through 316 of Ref. 3) is included in his reception model. We also determine the effect on system performance of simple channel-coding techniques. Our use of channel coding in spread-spectrum systems is similar to the case reported in Ref. 4, which involved frequency hopping. Both the (7, 4) Hamming code and (15, 7) BCH code are considered in our analyses. We assume hard decisions are made by the demodulator and that its error-producing mechanism results in independent error events. The latter assumption requires interleaving at the transmitter and de-interleaving at the receiver as a slowly fading channel model is considered. As the intended application is to packet transmission, interleaving does not present a severe system problem. A discussion of interleaving is given in Appendix 3A of Ref. 5.

The references to the channel-coding aspects of our study are important, as channel coding is found to be an effective method of combating multiuser interference in Spread-Spectrum Multiple Access (SSMA) systems. This was earlier found by Livine for no signal fading.[6] In an earlier study Turin[7] found that SSMA can tolerate considerably less multiuser interference in a fading channel than can be allowed in an additive white Gaussian noise channel. Adopting a Rayleigh fading model that seems less severe than the model used by Turin[7] for mobile communication applications, it is found that channel coding plus selection diversity performs well in a multiuser environment because the combination can be optimized. Channel coding used with selection diversity is found to perform better than selection diversity alone for the same spread-spectrum system bandwidth. In this sense it is both power and bandwidth efficient, as can be deduced from the similar study of Milstein et al.[4] in the absence of fading. This is true for the simple block codes mentioned above. Using more powerful codes and/or soft-decision decoding would give even greater gains in performance. Our approach has been to adopt a simple

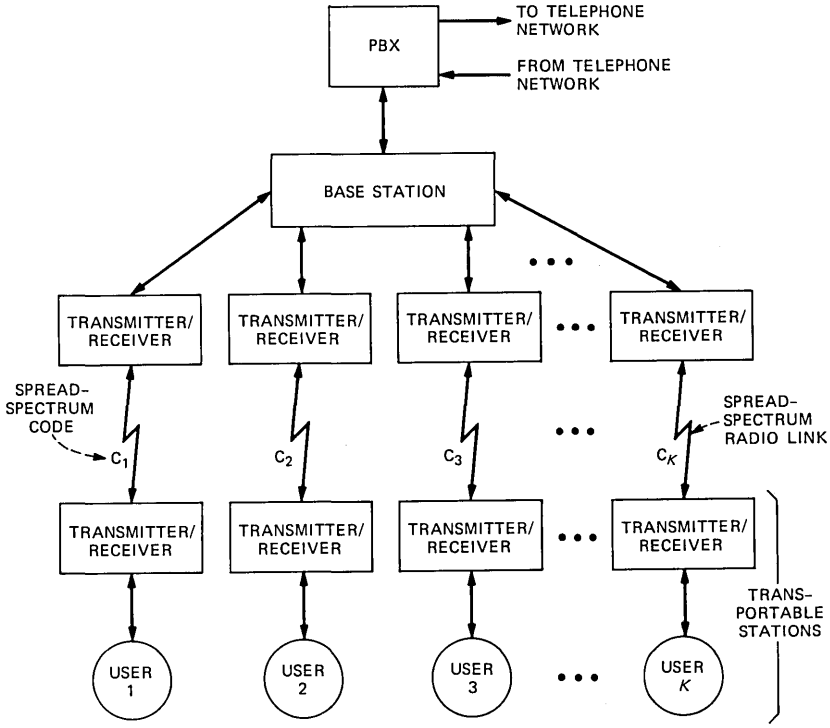


Fig. 1—A star-connected, indoor, wireless, local area network. Each user has a unique spread-spectrum code.

detection and decoding system to observe how a relatively simple system performs.

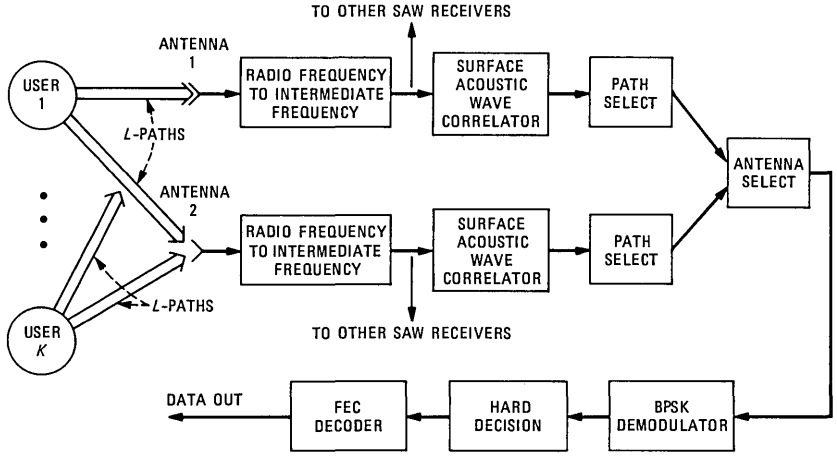
For this study we assume an indoor, wireless communication channel offering both voice and data services. The speech transmission rate for each user assumed in our system parameter study for IWC systems is 32 kb/s. Packet transmission is assumed and all users communicate through a central station in a star network architecture. Figure 1 is a simple block diagram of the system we analyze. Each active user in the system depicted in Fig. 1 has a unique spread-spectrum code, which is used for communication to the central station. The central station contains a bank of spread-spectrum receivers, one for each active user. Its function is to determine which subscribers are active and to detect the digital information sent in each case. The basis of the spread-spectrum receiver for each active user that exists in the central station is a Surface Acoustic Wave (SAW) device. Such devices have been found to be effective in such a role in the earlier study of Freret et al.[8] A tutorial on SAW devices can be found in Ref. 9. We note that the study in Ref. 8 proposed the use of spread-

spectrum diversity in IWC systems. This diversity is inherently supplied by the spread-spectrum modulation as long as the spread-spectrum bandwidth exceeds the coherence bandwidth of the slowly fading channel (see Ref. 10, page 480). More discussion on this point is presented in Section II. We note that spread-spectrum modulation can provide both asynchronous multiple access and also diversity reception.[7,10] Other advantages of spread-spectrum in IWC systems are discussed by Kavehrad.[1]

We present an analysis of the bit error probability for the link between any active user and its receiver in the central station. This link is shown in Fig. 1. As such, we are only considering the communication upstream from an active user to the central station. The downstream communication path is much simpler and will not be considered. We assume that average power control is used by all active users in that, on the average, all active user signals are assumed to arrive at the central station with the same power (in the upstream communication mode). The average here includes the fading statistics. Thus, the power control that must be used by each active user just depends on the distance and power law exponent for the link from a user to the central station and also on the static, shadow fading that is encountered. The sources of signal fading are nicely summarized in Section II of Ref. 11. We do not consider the dynamics of average power control in this paper.

The main contribution of the memorandum is to show that Kavehrad's analyses[1] can be extended to include selection diversity, and that this form of diversity can be used in conjunction with simple channel coding to give an SSMA system with an acceptable number of active users. For instance, for an IWC system having a multipath delay spread, T_m, of 100 nanoseconds, we find that a spread-spectrum code length of 255, a source data rate of 32 kb/s, and a (15, 7), double-error-correcting BCH code can support approximately 75 simultaneous users. If we envision a low-traffic office environment with a 10-percent channel utilization, a total of 750 subscriber terminals can be supported using the aforementioned method. This assumes that each code is shared among a group of subscribers in a contention mode of operation.

An outline of the paper is as follows. The system-fading and multipath model is described in Section II. Section III outlines the use of the computational technique from Ref. 1 to compute the average system error probability. Section IV considers an approximate computational technique based on a Gaussian assumption for the multiuser interference. Section V considers simple channel codes and Section VI contains our numerical results. Section VII presents an application of our computational results to two IWC scenarios for local area



BPSK – BINARY PHASE-SHIFT KEYING
FEC – FORWARD ERROR CORRECTING

Fig. 2—Per-user receiver for spread-spectrum, direct-sequence receiver using selection diversity.

networks, as well as our consideration of multipath outage as a performance criterion.

II. SYSTEM MODEL

2.1 Transmission model

Our mathematical model will depend heavily on the model developed by Kavehrad.[1] We will use the same notation and borrow heavily from Kavehrad's earlier analysis. Consider the block diagram of the multiuser, IWC channel shown in Fig. 2. The structure is exactly as in Fig. 1. However, the receiving systems for a single reference user, taken as user 1, are shown; for simplicity only a two-antenna system is depicted in Fig. 2.

In Fig. 1 each active user has a code waveform that consists of a periodic (period T) sequence of N, nonoverlapping rectangular waveforms (called chips), each of the duration T_c seconds. The length of the code waveform is T seconds, the reciprocal of the symbol transmission rate, where $T = NT_c$. The sequence of chip waveforms is the spread-spectrum code waveform, which for the kth user is denoted by $a_k(t)$. The data signal is binary with data symbols b_j^k, where the subscript denotes the jth time slot and the superscript denotes the data symbol for the kth user. If we let $P_T(t)$ denote a rectangular pulse of unit height and duration T, the transmitted signal for the kth user is

$$S_k(t) = Aa_k(t)b_k(t)\cos(\omega_c t + \theta_k) \tag{1a}$$

$$a_k(t) = \sum_{i=-\infty}^{\infty} a_i^k P_{T_c}(t - iT_c) \tag{1b}$$

and

$$b_k(t) = \sum_{j=-\infty}^{\infty} b_j^k P_T(t - jT), \tag{1c}$$

where a_i^k is the ith chip amplitude for the kth user, where $\omega_c T = 2\pi$ times an integer, ω_c is the carrier frequency in rad/s, A is the signal amplitude, and θ_k is the signal phase. In our analysis we shall have $k = 1, 2, \cdots, K$, where K will denote the number of simultaneous users.

In Fig. 2 we show L discrete multipath links between each user and each receive antenna at the central station. The low-pass equivalent impulse response of the passband channel for the link between the kth user transmitter and central station receiver is

$$h_k(\tau) = \sum_{\ell=1}^{L} \beta_{\ell k} \delta(\tau - \tau_{\ell k}) e^{j\Phi_{\ell k}}, \tag{2}$$

where for the kth user

$\beta_{\ell k} = \ell$th path gain

$\Phi_{\ell k} = \ell$th path phase

$j = \sqrt{-1}$

and

$\tau_{\ell k} = \ell$th path time delay.

We assume $\beta_{\ell k}$ is a Rayleigh random variable; $\Phi_{\ell k}$ is taken as uniform in $[0, 2\pi]$; and $\tau_{\ell k}$ is assumed uniform in $[0, T]$, where T is the data symbol interval. In the sequel the difference between the maximum and minimum values of $\tau_{\ell k}$ will be called the maximum multipath delay spread and will be denoted by T_m. Also, as a slowly fading channel is assumed, the variables in eq. (2) are assumed random but time-invariant.

The impulse response given in eq. (2) is characteristic of a discrete multipath channel and has the same functional form as that given in Ref. 12. The question is, How do we determine L in terms of communication system parameters? The basic result on the time resolution of signals using spread-spectrum signals is given in Section 1.5.3 of the recent textbook by Simon et al.[5] As one would expect, two signals must be separated by one chip time, T_c, in order to be resolved. We illustrate this in Figs. 3 and 4. Figure 3 is a system block diagram plus a diagram depicting the discrete multipath model. Figure 4 shows the response due to $L = 3$ multipath components. We assume that the

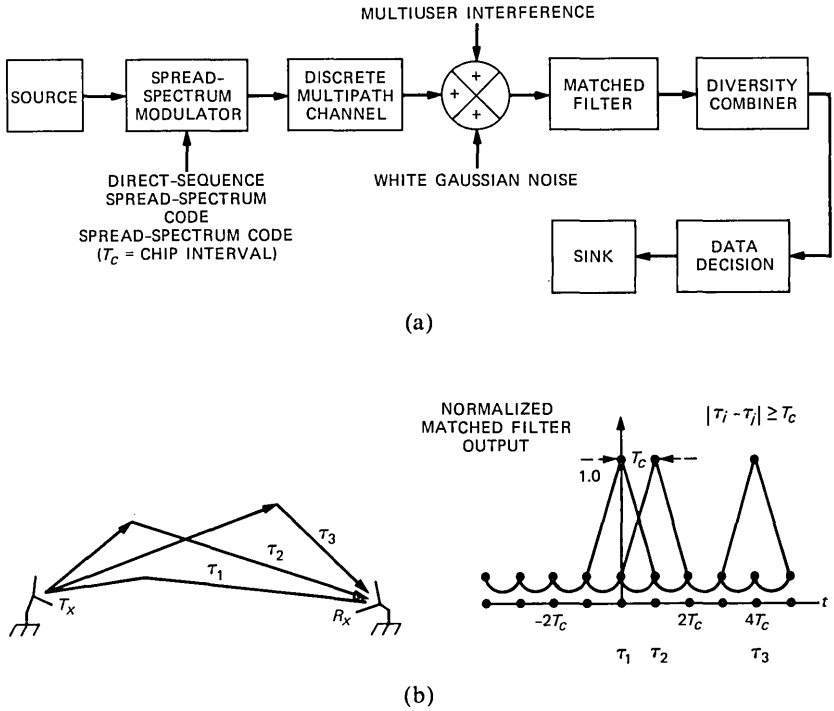


Fig. 3—Multiuser spread-spectrum system for (a) a baseband system and (b) a discrete multipath model.

maximum multipath delay spread, T_m, is less than T, the information bit interval, in order to avoid intersymbol interference. Using the result of time resolution of direct-sequence, spread-spectrum signals given above, one sees that

$$L = \left\lfloor \frac{T_m}{T_c} \right\rfloor + 1 = \lfloor T_m \cdot B_{ss} \rfloor + 1 \tag{3}$$

is the maximum number of resolved paths for a maximum multipath delay spread of T_m seconds. Also, in eq. (3) $\lfloor x \rfloor$ is the largest integer that is less than or equal to x and $B_{ss} = NR_0$, the one-sided bandwidth of the spread-spectrum signal, where $R_0 = T^{-1}$ and $T/T_c = N$ is the sequence length. Copies of the transmitted signal that arrive at unresolvable time differences are assumed to combine to give rise to the Rayleigh path gain of eq. (2). As such, we should assume that the time difference, $\tau_j - \tau_k$, is greater than T_c, where each individual τ is uniform in $(0, T)$. We take $\tau_j - \tau_k > 0$, which approximately is true as $T_c = T/N$ is small relative to T.

Actually, we feel that L in eq. (3) represents the maximum number

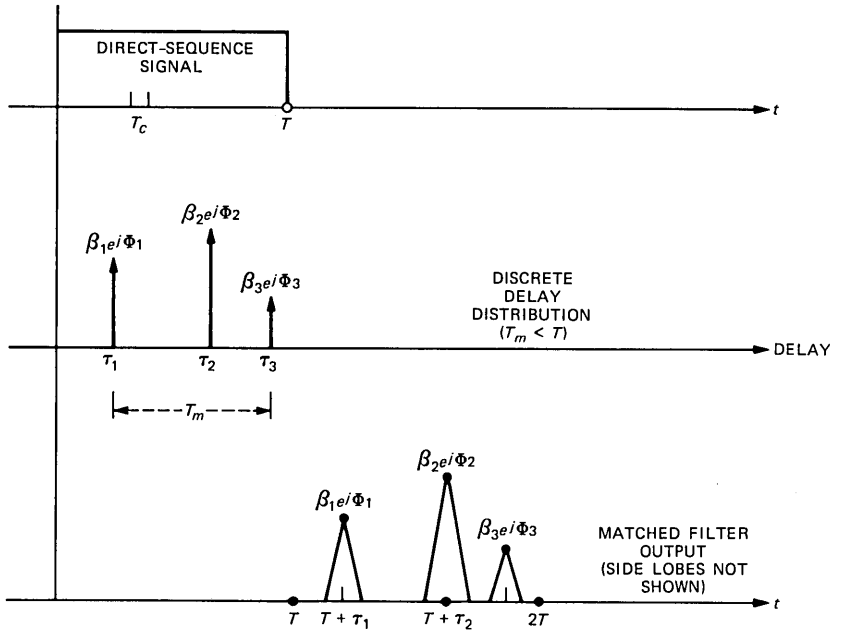


Fig. 4—Signal transmission model for a direct-sequence, spread-spectrum system and a discrete multipath model.

of approximately uncorrelated terms one could have in eq. (2). This is because L is based on the minimum resolution of direct-sequence, spread-spectrum signals. A random model is one in which L varies between unity and the maximum value given in eq. (3).

Our approach to treating L in the paper will be as follows. Up to the IWC parameter study presented in Section VII, we determine system performance in general for any K, L, and M. Here K is the number of simultaneous users, L the number of paths, and M the order of diversity. In Section VII we then adopt two models for L. In one, L is given by eq. (3). In the other, L varies uniformly between unity and the maximum value given in eq. (3). The results turn out to be insensitive to the model used for L. As better models for L evolve, our general results can be used to estimate performance in such cases.

If we combine eqs. (1a) and (2) and use the convolution integral, the received signal at the central station, which will be denoted as $r(t)$, is given by

$$r(t) = \mathrm{Re}\left\{\sum_{k=1}^{K}\int_{-\infty}^{\infty} h_k(\tau)\tilde{S}_k(t-\tau)\exp(j\omega_c t)d\tau\right\} + n(t), \qquad (4)$$

where $\tilde{S}_k(t)$ is the complex envelope of $S_k(t)$ for $\theta_k = 0$ and $\mathrm{Re}\{\cdot\}$

denotes the real part of a complex number. Upon use of eqs. (1), (2), and (4), we have

$$r(t) = A \sum_{\ell=1}^{L} \beta_{\ell 1} a_1(t - \tau_{\ell 1}) b_1(t - \tau_{\ell 1}) \cos(\omega_c t + \Phi_{\ell 1}) + A \sum_{\ell=1}^{L} \sum_{k=2}^{K} \beta_{\ell k} a_k(t - \tau_{\ell k}) b_k(t - \tau_{\ell k}) \cos(\omega_c t + \Phi_{\ell k}) + n(t). \quad (5)$$

The white Gaussian noise, $n(t)$, in eq. (5) will have a spectral height of $N_0/2$ W/Hz. In eq. (5) $\tau_{\ell k}$ will be uniform in $[0, T]$, $\Phi_{\ell k}$ uniform in $[0, 2\pi]$, and $\beta_{\ell k}$ will have the Rayleigh probability density function (pdf)

$$f_\beta(x) = \frac{x}{\rho_0} \exp\left(\frac{-x^2}{2\rho_0}\right) u(x), \quad (6)$$

where $u(x)$ is the unit step function, $u(x) = 1$ for $x \geq 0$ and zero elsewhere. As such, the average received signal-to-white-Gaussian-noise ratio is

$$\gamma_0 = E(\beta_{j1}^2) \frac{E_b}{N_0} = 2\rho_0 E_b / N_0, \quad (7)$$

where $E_b = A^2 T/2$, the signal energy per bit, and $E(\beta_{j1}^2) = 2\rho_0$, where for user one β_{j1} is the random gain of the jth signal path. In fact $\gamma_0 = E(\gamma)$, where $\gamma = \beta_{j1}^2 E_b/N_0$ has the exponential pdf

$$f_\gamma(y) = \gamma_0^{-1} \exp\left(\frac{-y}{\gamma_0}\right) u(y) \quad (8)$$

with γ_0 as in eq. (7).

The specification of our channel model is now complete. Note that the formulation represented by eq. (5) pertains to only the discrete multipath model whose impulse response is given by eq. (2). We note that the transmission model is similar, except for the specification of the fading parameters, to that used by Pursley.[13] In Ref. 14 multipath diversity reception is considered. However, a Gaussian assumption is used in the performance computations.

2.2 Receiver model

The input to the receiver for the reference user is given by the right-hand side of eq. (5). The first term in this equation represents all the copies of the transmitted, spread-spectrum signal that are available for detection. Let us assume that the receiver can ideally lock on to the term at delay τ_{j1} and phase Φ_{j1}. Then the jth decision variable for detection is given by

$$\xi_j = \int_0^T r(t)a_1(t - \tau_{j1})\cos(\omega_c t + \Phi_{j1})dt. \tag{9}$$

If the order of diversity used in the receiver is M, we will have ξ_j, $j = 1, 2, \cdots, M$ decision variables available for detection purposes. Substituting eq. (5) into eq. (9) yields the result

$$\begin{aligned}\xi_j = b_0^1 \frac{AT}{2}\beta_{j1} &+ \frac{A}{2}\sum_{\substack{\ell=1\\ \ell\neq j}}^{L}\beta_{\ell 1}\cos[\Phi_{\ell 1} - \Phi_{j1}] \\ &\cdot\int_0^T a_1(t - \tau_{\ell 1})b_1(t - \tau_{\ell 1})a_1(t - \tau_{j1})dt \\ &+ \frac{A}{2}\sum_{\ell=1}^{L}\sum_{k=2}^{K}\beta_{\ell k}\cos[\Phi_{\ell k} - \Phi_{j1}] \\ &\cdot\int_0^T a(t - \tau_{\ell k})b_k(t - \tau_{\ell k})a_1(t - \tau_{j1})dt \\ &+ \int_0^T n(t)a_1(t - \tau_{j1})\cos[\omega_c t + \Phi_{j1}]dt, \end{aligned}\tag{10}$$

where b_0^1 is the data bit to be detected. If one consults Ref. 1, it is clear that our eq. (10) is equivalent to eq. (8) of that reference with one difference; that is, in our eq. (10), L fading paths are assumed for each interferer, as it is important in this work. Following this same reference, one can express eq. (10) in the form

$$\begin{aligned}\xi_j = \beta_{j1}\frac{AT}{2}b_0^1 &+ \frac{A}{2}\sum_{\substack{\ell=1\\ \ell\neq j}}^{L}\beta_{\ell 1}\cos[\Phi_{\ell 1} - \Phi_{j1}] \\ &\cdot[b_{-1}^1 R_{11}(t_{\ell 1}) + b_0^1\hat{R}_{11}(t_{\ell 1})] \\ &+ \frac{A}{2}\sum_{\ell=1}^{L}\sum_{k=2}^{K}\beta_{\ell k}\cos[\Phi_{\ell k} - \Phi_{j1}] \\ &\cdot[b_{-1}^k R_{k1}(t_{\ell k}) + b_0^k\hat{R}_{k1}(t_{\ell k})] + \nu, \end{aligned}\tag{10a}$$

where $t_{\ell k} = \tau_{\ell k} - \tau_{j1}$, ν is Gaussian with zero mean and variance $N_0T/4$,

$$R_{k1}(\tau) = \int_0^\tau a_k(t - \tau)a_1(t)dt, \tag{10b}$$

and

$$\hat{R}_{k1}(\tau) = \int^T a_k(t - \tau)a_1(t)dt. \tag{10c}$$

Although eq. (10) is rather long, each term in this equation can be interpreted with respect to the block diagram shown in Fig. 2. The first term in eq. (10) represents the desired signal to be detected. The second term in eq. (10) is the self-interference for the reference user (say, User 1 in Fig. 2) due to sidelobes of the autocorrelation function of the spread-spectrum code of User 1. The third term in eq. (10) is the $L(K-1)$ multiuser interference terms from the $K-1$ other simultaneous users of the system. Finally, the last term in eq. (10) is the Gaussian random variable due to additive white Gaussian noise. Note that there are $L-1$ self-interference plus $L(K-1)$ multiuser interference terms in eq. (10). Thus, the total number of interference terms is $\eta = L(K-1) + L - 1 = LK - 1 = LK$ for a large LK. We will find in our computations that the per-user average error probability is, for all practical purposes, a function of $\eta = LK$. We note that ξ_j in eq. (10a) could correspond to any diversity term for the receiving system shown in Fig. 2. For instance, if there are two antennas and $L = 2$ (see Fig. 2), we have two antennas and two paths per antenna to give a total order of diversity of $M = 4$.

As we have noted previously, our eq. (10) is equivalent to eq. (8) of Ref. 1. Also eq. (10a) is equivalent to eq. (10) of this reference. However, in our detection procedure we assume that selection diversity is used. That is, we can find ξ_j in eq. (10) such that β_1 is the largest path gain relative to User 1, so that

$$\beta_1 = \text{Max}(\beta_{11}, \beta_{21}, \cdots, \beta_{M1}).$$

Let ξ be that value of ξ_j corresponding to path gain β_1. Then our eq. (10) with ξ_j replaced by ξ is equivalent to eq. (8) of Ref. 1, except that β_1 has the pdf of the maximum of the path gains $\beta_{j1}, j = 1, 2, \cdots, M$. As shown by Jakes,[3] pages 313 through 316, or Papoulis,[15] pages 139 and 140, the pdf of β_1^2, the maximum of the β_{j1}^2, is easily found. It is this pdf that will be needed in our error probability analyses.

III. ERROR PROBABILITY

We will return to the subject of the pdf of β_1^2 later. Recall that our eq. (10) is equivalent to eq. (8) of Ref. 1. If we mimic the development through to eq. (22) of this reference, we find that the probability of error, conditioned on β_1^2, the self-interference, and the multiuser interference is given by

$$P(e \mid \beta_1, x, z) = \frac{1}{2} \operatorname{erfc} \left\{ \sqrt{\frac{\beta_1^2 E_b}{N_0}} - \sqrt{\frac{E_b}{N_0}} (x + z) \right\}, \quad (11)$$

where

$$\operatorname{erfc}(u) = \frac{2}{\sqrt{\pi}} \int_u^\infty \exp(-y^2)dy$$

$$x = \frac{1}{T} \sum_{\substack{\ell=1 \\ \ell \neq j}}^{L} \beta_{\ell 1}\{b_{-1}^1 R_{11}(t_{\ell 1}) + \hat{R}_{11}(t_{\ell 1})\}\cos\nabla_\ell \tag{12}$$

$$z = \frac{1}{T} \sum_{\ell=1}^{L} \sum_{k=2}^{K} \beta_{\ell k}\{b_{-1}^k R_{k1}(t_{\ell k}) + b_0^k \hat{R}_{k1}(t_{\ell k})\}\cos\theta_{\ell k}, \tag{13}$$

where $t_{\ell k} = \tau_{\ell k} - \tau_{j1}$, $\nabla_\ell = \Phi_{\ell 1} - \Phi_{j1}$, and $\theta_{\ell k} = \Phi_{\ell k} - \Phi_{j1}$. The index j in, say, τ_{j1}, is taken as the delay for the path having the largest β_{j1}. The largest β_{j1} is denoted as β_1 and this random variable is independent of β_{i1}, $i \neq j$. Thus x and z, which are mutually independent, are also both independent of β_1.

Kavehrad's[1] technique of finding the average error probability, $P(e)$, was to integrate $P(e \mid \beta_1, x, z)$ with respect to the pdf of β_1^2 and then remove the conditioning on (x, z) using the method of moments. Actually our result for $P(e)$ in the case of selection diversity can be deduced from Kavehrad's mathematics. Let the pdf for β_1 be the Rayleigh pdf given in eq. (6). The pdf for β_1^2 is

$$f_{\beta_1^2}(y) = \frac{1}{\rho_0'} \exp\left\{-\frac{y}{\rho_0'}\right\} u(y) \tag{14}$$

with $\rho_0' = 2\rho_0 = E(\beta_1^2)$. In eq. (11) we have

$$P(e \mid x, z) = \int_0^\infty f_{\beta_1^2}(y)P(e \mid \beta_1, x, z)dy = K\left(\frac{E_b}{N_0}, \gamma_0, D\right), \tag{15}$$

where $\gamma_0 = \rho_0' E_b/N_0$, $D = x + z$,

$$K(v, \gamma_0, D) = \frac{1}{2}\operatorname{erfc}[-\sqrt{vD}] - \frac{1}{2}\sqrt{\frac{\gamma_0}{\gamma_0 + 1}} \exp\left[\frac{-vD^2}{\gamma_0 + 1}\right] \cdot \operatorname{erfc}\left[-\sqrt{\frac{\gamma_0}{\gamma_0 + 1}} D\sqrt{v}\right] \tag{16}$$

and $v = E_b/N_0$. This result is the same as in eq. (30) of Ref. 1.

It turns out that a result similar to eq. (14) is obtained when β_1^2 is the maximum of the β_{j1}^2, $j = 1, 2, \cdots, M$. If all the β_{j1}'s are Rayleigh with $E(\beta_{j1}^2) = 2\rho_0 = \rho_0'$,

$$f_{\beta_1^2}(y) = \frac{M}{\rho_0'}\left(1 - \exp\left[\frac{-y}{\rho_0'}\right]\right)^{M-1} \exp\left[\frac{-y}{\rho_0'}\right] u(y), \tag{17}$$

as follows from eq. (5.2-7) of Jakes.[3] Using the binomial theorem in eq. (17),

$$f_{\beta_1^2}(y) = M \sum_{k=0}^{M-1} \binom{M-1}{k} \frac{(-1)^k}{(k+1)\rho'_{0k}} \exp[-y/\rho'_{0k}]u(y), \tag{18}$$

where $\rho'_{0k} = \rho'_0/(k+1)$ and

$$\binom{N}{k} = \frac{N!}{(N-k)!k!}. \tag{19}$$

The evaluation of the integral in eq. (15) for the pdf in eq. (18) will just involve a summation of the $K(\cdot,\cdot,\cdot)$ function of eq. (16). This follows as the pdf in eq. (18) is just a summation of the exponential pdf's, $(\rho'_{0k})^{-1}\exp(-y/\rho'_{0k})$. Thus, use of eqs. (14) and (15) in the evaluation of the integral in eq. (15) for $f_{\beta_1^2}(y)$ in eq. (18) yields

$$P(e \mid x, z) = M \sum_{k=0}^{M-1} \binom{M-1}{k} \frac{(-1)^k}{k+1} K\left(\frac{E_b}{N_0}, \frac{\gamma_0}{k+1}, D\right), \tag{20}$$

where the $K(\cdot,\cdot,\cdot)$ function is as specified in eq. (16). This is the main mathematical result of this study. Kavehrad[1] showed how to average the $K(\cdot,\cdot, D)$ function with respect to $D = x + z$. To remove the dependence of $P(e \mid x, z)$ on $D = x + z$, one just carries out Kavehrad's procedure once, to get the moments of $D = x + z$, and then evaluates the resulting $K[\cdot, \gamma_0/(k+1), D]$ function for all $\gamma_0/(k+1)$. In mathematical terms,

$$P(e) = M \sum_{k=0}^{M-1} \binom{M-1}{k} \frac{(-1)^k}{k+1} \sum_j w_j K\left(\frac{E_b}{N_0}, \frac{\gamma_0}{k+1}, \zeta_j\right), \tag{21}$$

where (w_j, ζ_j) are the weights and nodes, respectively, of the Gauss-Quadrature algorithm (see Appendix C of Ref. 1). For $M = 1$ the result in eq. (21) reduces to eq. (32) of Kavehrad's analysis in Ref. 1.

IV. GAUSSIAN ASSUMPTION

The Gaussian assumption is to take all the multiuser interference as Gaussian noise. We will base our calculation of the average error probability on eq. (10a), which is equivalent to eq. (10) of Ref. 1. The last term in eq. (10a), ν, is a Gaussian random variable having zero mean and variance $N_0T/4$. The first term in eq. (10a) is the signal term and it has average power, $\beta_1^2A^2T^2/4$ for a fixed β_1. The rest of the terms in eq. (10a) are all mutually independent. To calculate the total power of this term, we must evaluate a term like

$$\epsilon^2 = E\left\{\frac{[\alpha_{-1}R(t_1) + \alpha_0\hat{R}(t_1)]^2}{T^2}\right\}, \tag{22}$$

where $\alpha_i = \pm 1$, $i = -1, 0$, are independent binary variables. In eq. (22), ϵ^2 was shown by Pursley[16] to have the value $2/(3N)$, where N is the sequence length of the Gold codes considered in Ref. 16.

There are approximately $\eta = LK$ such expectations in eq. (10a). Hence, for a fixed β_1 we have

$$\text{signal power} = \left(\frac{AT}{2}\right)^2 \beta_1^2$$

$$\text{interference power} = \eta \left(\frac{AT}{2}\right)^2 \epsilon^2 E(\beta^2)/2$$

and

$$\text{noise power} = N_0 T/4,$$

where β denotes a Rayleigh random variable and is any of the independent identically distributed random variables $\beta_{\ell k}$'s excluding β_1. The term $E(\beta^2)/2$ occurs in the interference power as $\beta \cos \theta$ is Gaussian with zero mean and variance $E(\beta^2)/2$ as θ is uniform in $[0, 2\pi]$. Thus, with the Gaussian assumption, the error probability conditioned on β_1 is given by

$$P(e \mid \beta_1) = \frac{1}{2} \operatorname{erfc}(\sqrt{\gamma}). \tag{23}$$

Note that in eq. (23) γ is equal to half the signal-to-noise plus interference power ratio; hence

$$\gamma = \frac{\beta_1^2 E_b}{(LK)\epsilon^2 E_b E(\beta^2) + N_0} \tag{24}$$

with $\epsilon^2 = 2/(3N)$ given by eq. (22) and $E_b = A^2 T/2$. The average value of γ is

$$\gamma_0 = \frac{\overline{E_b}}{\eta \epsilon^2 \overline{E_b} + N_0}, \tag{25}$$

where $\overline{E_b} = E(\beta^2) E_b$ and $\eta = LK$. For $N_0 = 0$ we have $\gamma_0 = 3N/2\eta$, which is a result to be used in what follows.

When β_1^2 in eq. (24) has the pdf in eq. (14) with $E(\beta_1^2) = 2\rho_0 = \rho_0'$, Proakis[10] in his textbook shows that the average of eq. (23) with respect to β_1 is

$$P(e) = p(\gamma_0) = \frac{1}{2}\left\{1 - \sqrt{\frac{\gamma_0}{\gamma_0 + 1}}\right\}. \tag{26}$$

When selection diversity is used, the pdf for β_1^2 has the form in eq.

(18), which is just a summation of exponential pdf's. Accordingly, $P(e)$ for this case is just a sum of the terms in eq. (26), viz,

$$P(e) = M \sum_{k=0}^{M-1} \binom{M-1}{k} \frac{(-1)^k}{k+1} p\left(\frac{\gamma_0}{k+1}\right), \tag{27}$$

a result given by Sundberg.[17]

A more complicated, but higher performance, form of diversity is Maximal Ratio Combining (MRC). Here the gain and phase of each signal term must be known. These gains and phases are then used to coherently combine individual path terms to form a single decision variable for the data detection process. The decision statistic, assuming perfect coherence, is just the summation of $\beta_j \cdot \xi_j$, $j = 1, 2, \cdots, M$, where ξ_j is given in eq. (10). The error performance of this form of diversity can also be given in terms of γ_0 in eq. (25). If the interference terms are assumed to be uncorrelated from one diversity branch to another, then the result is given in eq. (7.4.15) of Proakis[10] text, viz,

$$P_e = [p(\gamma_0)]^M \sum_{k=0}^{M-1} \binom{M-1+k}{k} [1 - p(\gamma_0)]^k, \tag{28}$$

where $p(\gamma_0)$ is given in eq. (26).

We have not yet determined the error performance for MRC using the method of moments. Therefore, for this form of diversity we will have to rely on the Gaussian assumption. We will estimate the accuracy of the Gaussian assumption for the case of selection diversity and then apply this to the MRC case.

V. CHANNEL CODING

We shall be interested in the performance of two simple block channel codes used in conjunction with selection diversity. These are the (7, 4) Hamming code and the (15, 7) BCH code. The former corrects one channel error while the latter corrects two channel errors in a coded block. Such codes are discussed in the introductory textbooks by Pless[18] and Lin and Costello.[19]

For a channel code that corrects t-errors, the bit error probability is given in eq. (25) of Milstein et al.[4] as

$$P_{bt} = \frac{1}{n} \sum_{i=t+1}^{n} i \binom{n}{i} p_e^i (1 - p_e)^{n-i}, \tag{29}$$

where for simplicity we denote the channel error probability in eq. (21), $P(e)$, by p_e and n is the coded block length. This is an approximation and the assumption is made that channel errors are independent.

We have done some calculations with eq. (29) for the (7, 4) Hamming

code and the (15, 7) BCH code. However, the formulas given below are more precise, as they are based on the weight distribution of the codes used in our study. The two approaches never gave a difference in bit error rate of more than 66 percent. For the (7, 4) Hamming code we show in Appendix A that

$$P_{b1} = 9p_e^2(1 - p_e)^5 + 19p_e^3(1 - p_e)^4 \tag{30}$$

for small p_e. As this is a perfect code the result in eq. (30) is exact for small p_e. The result in eq. (29) gives $6p_e^2$ not $9p_e^2$ for the first term in eq. (30) when p_e is small.

For the (15, 7) BCH code we show in Appendix A that for small p_e,

$$P_{b2} = 150p_e^3(1 - p_e)^{12} + 512p_e^4(1 - p_e)^{11}. \tag{31}$$

Some approximations are involved here, as the code is not perfect. However, the weight distribution of the code is considered. The formula in eq. (29) gives $P_{b2} = 91p_e^3$ for small p_e.

VI. NUMERICAL RESULTS

The computer programs developed by Kavehrad[1] for the method of moments were modified to incorporate the moments of interference terms in the new form in eq. (10a). The new program was simply adapted to perform the computations needed to evaluate eqs. (20) and (21). Our computations will be for the Gold sequences of length 127 and the Kasami sequences of length 255. Initial loadings to generate these codes were taken from Ref. 20.

Before discussing our numerical results let us just review the main parameters of our model. They are

N = spread-spectrum sequence length

L = number of multipath links

M = number of terms used for diversity

K = number of simultaneous users

and

γ_0 = average signal-to-noise ratio.

In our computations we are interested in the case in which L is small, M is moderate, and K is large. For the most part, we will concentrate on the so-called[5] noise floor average error probability. This is the error probability when the thermal noise is absent; that is, when $N_0 = 0$. Our computations were most easily done for $1 \leq K \leq 15$ and $1 \leq L \leq 30$.

We will examine the three hypotheses listed below. The first one follows.

6.1 Error probability is approximately a function of only $\eta = KL$

Our hypothesis is that the error probability is approximately a function of only $\eta = KL$ and not of K and L themselves. This is certainly true in the case when a Gaussian assumption is made, as can be seen by an examination of eqs. (25), (26), (27), and (28) of Section IV. The goal of our system analysis is to estimate performance for, say, $K = 90$ and $L = 2$. We wish to do this by computing the error probability for $\eta = KL$, with, for instance, $K = 15$ and $L = 12$, which also gives $\eta = 180$.

The results of our computations for $N = 255$, the Kasami code, are shown in Table I. For error probabilities in the channel-coded cases, that is, P_{b1} and P_{b2}, of around 10^{-4} we see that, for the most part, we are in error at most by a factor of 2 or 3. This is an acceptable error factor for practical error probabilities. An error factor of 6 to 10 would not be acceptable in our view. This point takes us to our next hypothesis.

6.2 Coding plus selection diversity is power efficient

To evaluate the error performance when channel coding is used we take the result of the computation represented by eq. (21) and substitute it into (30) or (31). Equation (30) is for single-error correction and eq. (31) is for double-error correction.

We will compare channel coding plus selection diversity of order M versus selection diversity alone of order $2M$. This will be done for the single-error-correcting (7, 4) Hamming code. Thus P_{b1} is given in eq. (30). For large γ_0 it is easily shown that P_e in eq. (27), for selection diversity and a Gaussian assumption, is given by $c\gamma_0^{-M}$, where c is a constant. For various constants, c, see Table 1 of Ref. 17. Thus, using the (7, 4) Hamming code plus diversity takes $M \rightarrow 2M$, since the channel error probability is squared to give the decoded error proba-

Table I—The data to test the $P_e = f(\eta)$, $\eta = LK$, hypothesis for $N = 255$

M	η	K	L	P_e	P_{b1}	P_{b2}
4	60	10	6	0.41E-02	0.15E-03	0.98E-05
4	60	6	10	0.50E-02	0.22E-03	0.18E-04
6	90	15	6	0.85E-03	0.65E-05	0.91E-07
6	90	6	15	0.19E-02	0.32E-04	0.10E-05
6	150	15	10	0.41E-02	0.15E-03	0.98E-05
6	150	10	15	0.50E-02	0.22E-03	0.18E-04
6	180	15	12	0.69E-02	0.41E-03	0.45E-04
6	180	12	15	0.91E-02	0.57E-03	0.72E-04
8	180	15	12	0.35E-02	0.11E-03	0.62E-05
8	180	10	18	0.44E-02	0.17E-03	0.12E-04
8	240	15	16	0.88E-02	0.67E-03	0.91E-04
8	240	12	20	0.97E-02	0.80E-03	0.12E-03

bility and thus doubles the order of diversity. One more point is important in our comparison. Diversity and channel coding are similar in nature, since both try to exploit the redundancy in the transmitted signal. Our codes require bandwidth expansion to achieve this redundancy, but space diversity does not. Now the number of discrete multipath links between transmitter and receiver is given by eq. (3). Since the signal bandwidth, B_{ss}, is smaller for selection diversity alone, we evaluate its performance when L is replaced by $L - 1$ for $L = 2$ or 3, where the channel-coded system is taken to produce L multipath terms. Of course, for larger L's, the former should be replaced by $L - 2$, and so on. In any case, the uncoded system is subject to less interference than the coded system. The results of our computations are shown in Fig. 5. Also shown in Fig. 5 are two isolated points for

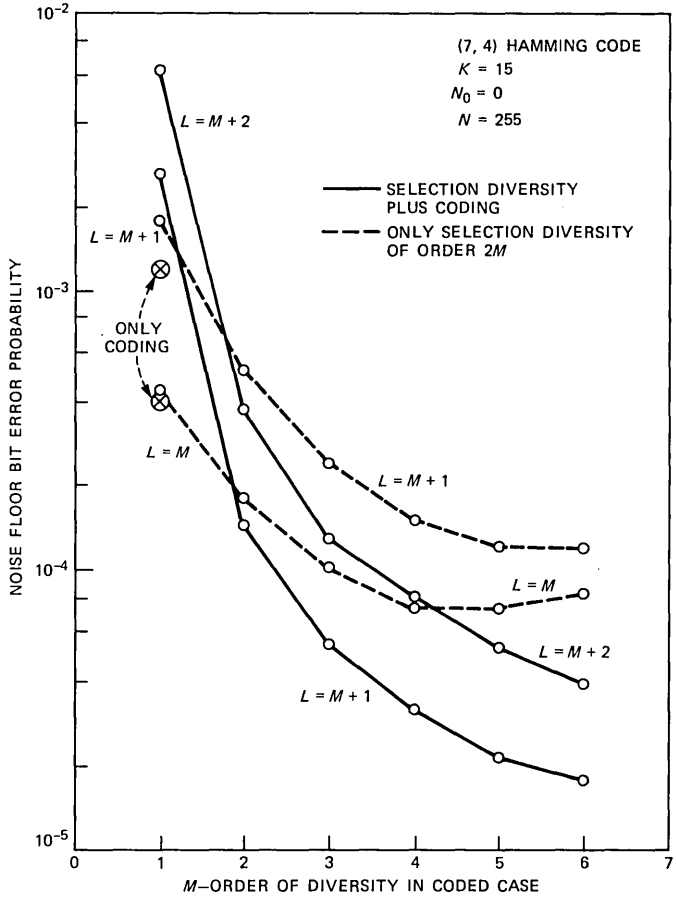


Fig. 5—Comparison of selection diversity plus coding versus selection diversity alone.

coding alone, that is, $M = 1$. One notes that, except for low M, the combination of channel coding plus selection diversity is significantly better than using selection diversity alone.

We note that selection diversity alone can saturate to a poor value of error probability as M gets large. From Jakes[3] it is well known that the signal-to-noise ratio (s/n) performance of selection diversity becomes poor relative to MRC as M grows. In the case of multipath diversity, performance becomes even worse. To see this point let us first consider MRC. Jakes[3] shows that, for antenna diversity of order M, the average s/n is $M\gamma_0$ (see page 3.19 of Ref. 3), where γ_0 is the average s/n for no diversity. Let the only source of noise for multipath diversity be the self-noise term in eq. (10). Accordingly, for multipath diversity alone γ_0 must be changed to γ_0/M and the average s/n with MRC diversity is γ_0 for all M. Thus, multipath diversity allows for no improvement in average s/n. In general, let there be L_S antennas and L_M multipath diversity terms giving $M = L_S \cdot L_M$. In the sequel we always take $L_M = L$, the number of paths in our multipath model. Then the average s/n is $L_S\gamma_0$ and the improvement is only due to L_S. With selection diversity Jakes[3] shows that the average s/n is

$$E(\gamma) = \gamma_0 \sum_{k=1}^{M} \frac{1}{k}.$$

For $M = 2$ we have $E(\gamma) = 0.75\ \gamma_0$ as for multipath diversity alone, since we must replace γ_0 by $\gamma_0/2$ due to self-interference. For $M = 4$ with $L_S = L_M = 2$ we have $E(\gamma) = 2\gamma_0$ for MRC and we have $E(\gamma) = 25\gamma_0/24$ for selection diversity, a loss of about 3 dB to MRC. Fortunately, the system error probability is not a function of $E(\gamma)$ alone, as it is a polynomial as γ_0^{-1}. In any case, for acceptable performance we will find that selection combining requires both antenna and multipath diversity. However, this may not be necessary for MRC or equal gain combining; only multipath diversity may suffice.

6.3 Coding plus selection diversity is power and bandwidth efficient

Our next comparison involves the performance of coding plus selection diversity versus selection diversity alone for the same system bandwidth. We did this by finding the selection diversity performance of the 127-length code (the Gold codes) with the (7, 4) Hamming code versus the length 255 code (the Kasami codes) with no channel coding. The result is plotted in Fig. 6. As Milstein et al.[4] found earlier, simple error-correcting codes are an effective way to improve the performance of spread-spectrum systems for the same system bandwidth.

6.4 Performance: coding plus selection diversity

The results of our computations using the method of moments are

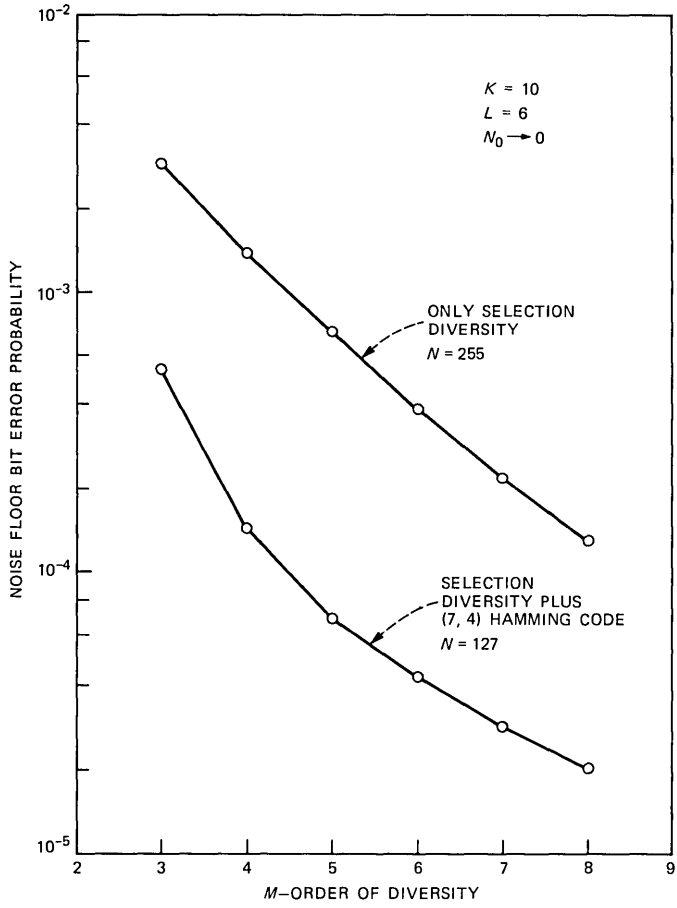


Fig. 6—Comparison of coding versus no coding in the same spread-spectrum bandwidth.

given in Figs. 7, 8, and 9. Figures 7 and 8 depict the noise floor error probability as $N_0 = 0$. Figure 7 is for small L and Fig. 8 is for large L. These computations are for independent values of K, L, and M; M is not a function of L, as will be the case in Section VII. We are interested in large L and $P_b = 10^{-4}$ in our system analysis, which will be discussed in the next section, where extensive use will be made of the results in Figs. 7 and 8. Figure 9 presents the performance with $N_0 \neq 0$, which will not be used in the sequel, since for IWC systems analysis is completely based on the noise floor error probability.

6.5 Computations: Gaussian assumption

It is clear that computation of the system error probability when a Gaussian assumption is invoked is quite simple. This is true for either

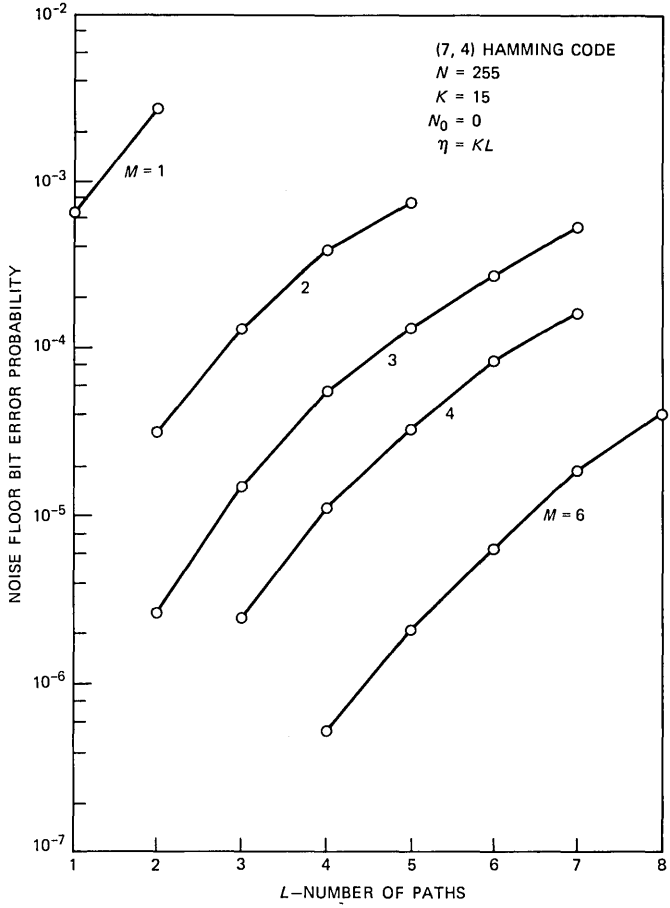


Fig. 7—Noise floor error probability for the (7, 4) Hamming code and various orders of diversity.

selection or MRC diversity, as can be observed by referring to eqs. (27) and (28), respectively. Since the method of moments precisely computes the system error probability, we can assess the goodness of the Gaussian approximation.

Our calculations are plotted in Figs. 10a and b. The Gaussian approximation underestimates the noise floor error probability. This gets worse as the number of interferers increases, a counterintuitive result based on intuition related to the central limit theorem. Around a 10^{-4} error probability, however, the discrepancy is acceptable. Because the Gaussian assumption underestimates the error probability, it will, for a fixed error probability, lead to an overestimation in the number of simultaneous users of a spread-spectrum multiple access system. We present the degree of this overestimation in Table II. In

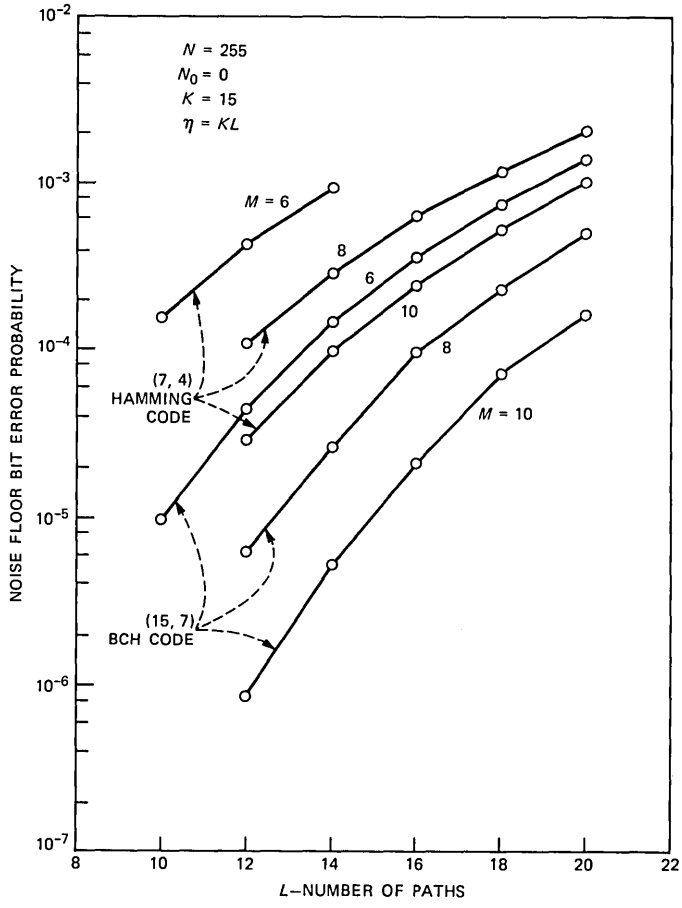


Fig. 8—Noise floor error probability for the (7, 4) Hamming code and (15, 7) BCH code for various orders of diversity.

this table both the Hamming and BCH codes are considered for two cases of diversity, $M = 4$ and 6. In both cases we are interested in that value of $\eta = KL$ that gives a noise floor error probability of approximately 10^{-4}. To do this we consider the Gold code of length 127 and the Kasami code of length 255. In the case of a Gaussian assumption the system performance depends only on $\gamma_0 = 3N/(2\eta)$. Thus, if N is doubled, so should η for the same value of γ_0. The percent error in this assumption is shown in Table II. For the orders of diversity of interest, the error is at most 20 percent.

We have also done computations for MRC by invoking the Gaussian assumption. This procedure uses eq. (28) and just depends on the parameter, $\gamma_0 = 3N/(2\eta)$, $\eta = KL$. We then reduce the value of η produced by this computation by 20 percent for, say, $P_e = 10^{-4}$, as this

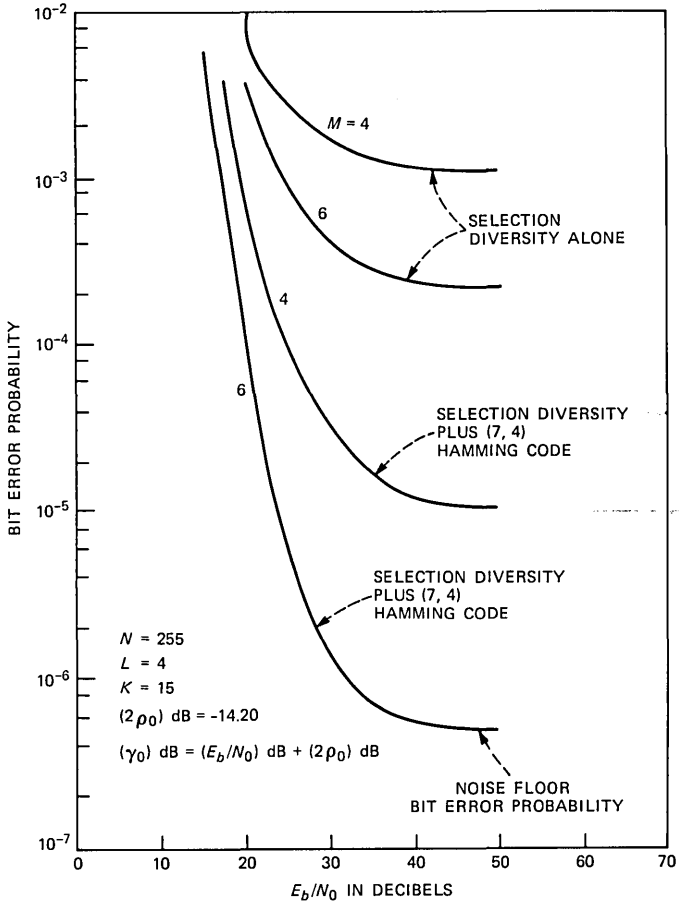


Fig. 9—Bit error probability as a function of E_b/N_0 for $M = 4$ and $M = 6$.

was the error in the case for selection diversity. The results of a limited number of such computations will be discussed in the next section.

VII. IWC PARAMETER STUDY

7.1 T_m = 100 nanoseconds

Measurements by Saleh and Valenzuela[21] have established the multipath delay spread in the Crawford Hill building at AT&T Bell Laboratories, Holmdel, New Jersey. The measurements indicate that the maximum delay spread is usually $T_m = 100$ ns. The distance over which these measurements were taken was approximately 300 ft.

The application that interests us is for 32-kb/s digital speech. We take this as the source rate. The service supplied also will include a 9.6-kb/s data service, and such a source would be channel coded up to

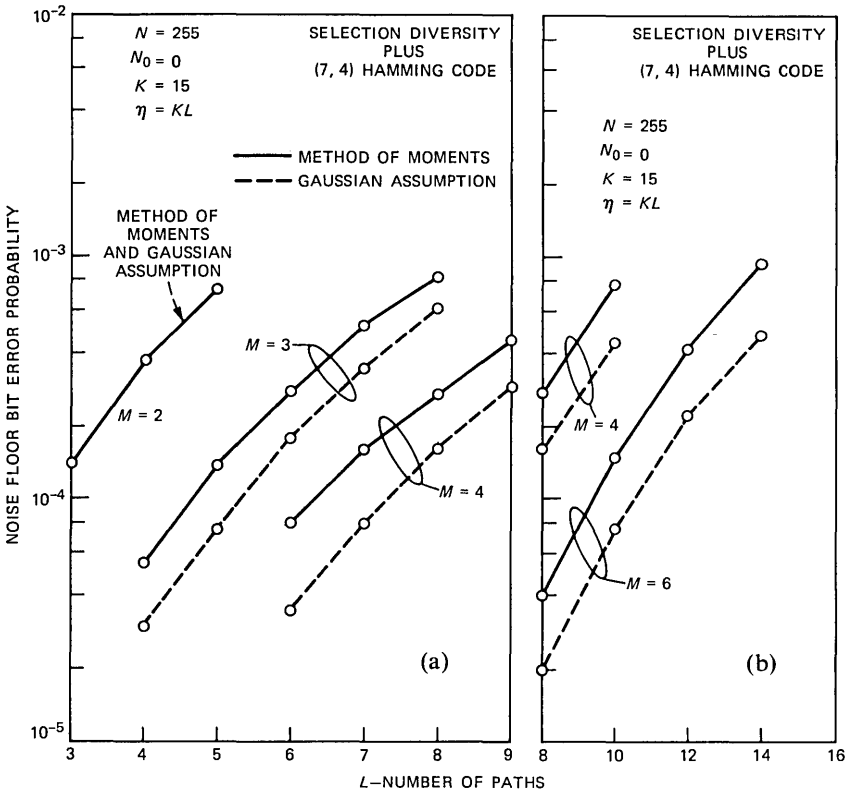


Fig. 10—The Gaussian assumption for (a) low orders of diversity and (b) high orders of diversity.

Table II—Error in the Gaussian assumption when the sequence length is doubled. The calculations of η are for $P_b \approx 10^{-4}$ and $\eta = KL$.

M	Code	η 127	η 255	η Gaussian 255	Percent Error
4	(7, 4)	60	105	120	14.3
4	(15, 7)	90	150	180	20.0
6	(7, 4)	80	150	160	6.7
6	(15, 7)	120	210	240	14.3

the 32-kb/s rate to provide the extra error protection needed for data. We focus on a threshold average error rate of 10^{-4} for speech. For 1000 bit packets this translates into approximately a 10-percent packet error probability. Valenzuela[22] and Wong et al.[23] have developed interpolation schemes to handle such packet loss rates. In any case, for a 32-kb/s source rate, the bandwidths of the various spread-spectrum

Table III—Transmission parameters for different sequence lengths

N	No Coding	(7, 4) Code	(15, 7) Code
(a) Bandwidth in MHz of various spread-spectrum sequence lengths for a source rate of 32 kb/s			
127	4.06	7.11	8.70
255	8.16	14.20	17.50
512	16.40	28.67	35.11
(b) Number of discrete multipath components when $T_m = 100$ ns			
127	1	1	1
255	1	2	2
512	2	3	4
(c) Number of discrete multipath components when $T_m = 250$ ns			
127	2	2	3
255	3	4	5
512	5	8	9

systems that we will consider are given in Table IIIa. The IWC application here presumes overlay signaling,[24] where spread-spectrum users' signal coexists with that of users of other services in a lightly loaded part of the radio frequency band.

A crucial parameter in our study will be the number of discrete paths, L, for a given maximum multipath delay spread and spread-spectrum bandwidth as predicted by eq. (3). Thus, for $N = 255$ in Table III and for $T_m = 100$ ns, we let $L = 2$. Other values of L are given in Table IIIb. From Fig. 8 we note that for the (15, 7) BCH code (double-error-correcting), we can have $L = 14$ for $K = 15$ in order that $P_b \approx 10^{-4}$ when $M = 6$. As $K = 15$ we have $\eta = LK = 210$. We now invoke our assumption about the fact that the system error rate is, to a good practical approximation, just a function of $\eta = KL$. Thus, if $L = 2$ we can get $K = 105$ simultaneous users since $\eta = 210$. The order of diversity $M = L_S \cdot L$, where L_S is the number of antennas and L the discrete order of spread-spectrum diversity. Therefore, $L_S = 3$ antennas at the central station is needed to support 105 simultaneous users.

Let us see now how many simultaneous users the single-error-correction system can support. For $M = 6$ from Fig. 8 we get $L = 10$ and thus $\eta = 150$. As $L = 2$, our assumption $P_e \approx f(KL)$ gives $K = 75$ active users. This is for $L_S = 3$ antennas.

With the sort of computation we have just outlined—through use of Fig. 8—and other related calculations, we can construct Table IV. The bandwidth efficiency measure given in Table IV will be discussed below. Also given are some performance estimates for the MRC form of diversity with no channel coding. When the spread-spectrum se-

Table IV—The number of simultaneous users in terms of sequence length, N, and L_s. The number of users is given in columns 3, 4, and 5. The order of diversity is $L_s \cdot L$, where L is given in Table IIIb. We have set $T_m = 100$ ns, $P_b \approx 10^{-4}$.

N	L_s	(7, 4) Code	(15, 7) Code	MRC at $2N$	Bandwidth Efficiency (15, 7) Code	Bandwidth Efficiency MRC at $2N$
127	2	20	40	6	0.14	0.02
255	1	23	38	6	0.07	0.01
255	2	50	75	60	0.12	0.12
255	3	75	105	132	0.19	0.25
512	1	42	60	—	0.05	—
512	2	80	108	—	0.10	—

quence length, N, is, say, 255, we have set $N = 512$ for MRC in order that the MRC system and selection diversity with channel coding system occupy the same bandwidth. The estimates for MRC were computed using the Gaussian assumption, $\gamma_0 = 3N/(2\eta)$, and eqs. (26) and (28). Such estimates of K were then reduced by 20 percent in accordance with our earlier results on the Gaussian approximation. We have made this reduction in computing the data of Table IV. Subject to such estimates we note that $M = 6$ is needed for MRC to outperform the combination of selection diversity and channel coding.

We note that the estimate for K for an $N = 512$ sequence length was determined as follows. The value of $\eta = KL$ for $N = 255$ was doubled and the result was reduced by 20 percent. This is in keeping with our findings regarding the Gaussian assumption (see Table II).

In Tables Va and b we present results when L is random. We let L vary from unity up to the maximum value, $L_{\max}$, given by the right-hand side of eq. (3). Each value of L is taken to occur with a probability of $1/L_{\max}$. We display the average value of K and also the value K corresponding to $L = L_{\max}$ in Tables Va and b. Actually, the value of K for each value of L, $1 \le L \le L_{\max}$, differed only slightly from the average value of K. We find this invariance because as L decreases, the order of diversity, $M = L_s L$, decreases, but so does the maximum number of interference terms, $\eta = KL$, thus giving rise to approximately a fixed P_b. Note that we cannot have only one antenna in the random model, since when $L = 1$ all diversity is lost.

In Fig. 11 we have plotted the bandwidth efficiency of some of our schemes. This is given by

$$\text{BE} = \frac{K \cdot (\text{Code Rate})}{N},$$

where N is the spread-spectrum code length and K is the number of

Table V—A comparison of the estimate of the number of simultaneous users when $L = L_{max}$, as given by eq. (3), and when $L = 1, 2, \cdots, L_{max}$, each with probability $1/L_{max}$. Average K is over such L.

N	L_s	L_{max}	K for L_{max}	Average K for $1 \le L \le L_{max}$
(a) Data are for the (7, 4) code with $T_m = 100$ ns				
127	2	1	20	20
255	2	2	50	50
255	3	2	75	75
512	2	3	80	79
(b) A comparison of the estimate of the number of simultaneous users for the (15, 7) code				
127	2	1	40	40
255	2	2	75	75
255	3	2	105	112
512	2	4	108	115

simultaneous users as estimated using the procedure just described above. We note that the double-error-correcting system is 11 percent more bandwidth efficient than the single-error-correcting system. Tabular data on bandwidth efficiency are given in Table IV. Although the efficiency values in Fig. 11 are rather low, we remember that in overlay signaling the values represent the order of frequency reuse.

We have also placed other points on our bandwidth efficiency plot in Fig. 11. These points are for less severe channel models than the one we consider. If we assume an Additive White Gaussian Noise (AWGN) channel, the performance follows from eq. (24). In eq. (24) let $N_0 = 0$ so that we get the noise floor error probability and assume that the multiuser interference follows the Gaussian model. Furthermore, let $b = \beta_1^2/E(\beta_1^2)$ and $\eta = LK$. Combining eqs. (23) and (24) for the bandwidth efficiency, there follows

$$\mathrm{BE} = \frac{3b}{2L\{\mathrm{erfc}^{-1}(2P_e)\}^2}. \tag{32}$$

In eq. (32) for $P_e = P_b = 10^{-4}$ we have BE = 0.22b/L. For the AWGN channel $\beta_1 = \beta = 1$ and $L = 1$ to give BE = 0.22, which agrees with Turin's[7] result (see Table 1 of Ref. 7; our result is slightly higher, as we use coherent binary phase-shift keying rather than differential binary phase-shift-keying modulation). Thus, for the same bandwidth as used in Table IV, as we have $\eta = 0.22$, $N = 512$ gives $K = 112$ and for no coding or diversity.

Another case of interest is when the signal term has a deterministic

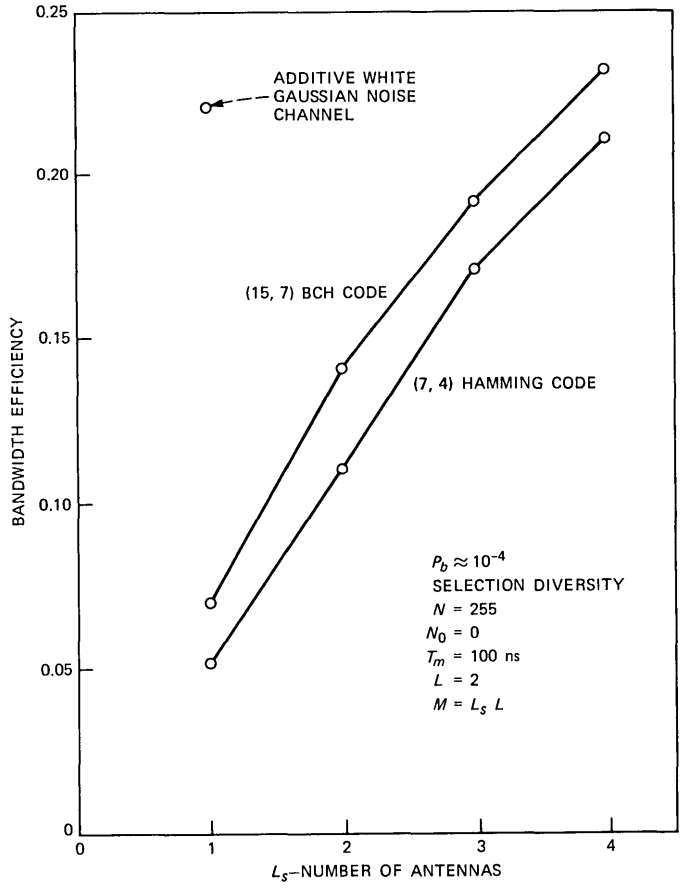


Fig. 11—Bandwidth efficiency of error-correction-coded spread-spectrum systems. For $N = 255$, selection diversity alone requires $L_s = 8$ to get an efficiency of 0.11.

gain but the multiuser interference is subject to Rayleigh fading (see Case 1, Section 4.2 of Ref. 1). As stated in Ref. 1, this is the case that the reference transmitter is stationary and there is not much movement in the indoor medium. For $L = 2$, that is, $T_m = 100$ ns, $N = 255$ and $R_0 = 32$ kb/s, we have BE $= 0.11b$, where $b = \beta_1^2/E(\beta^2)$. If $b = 2$, meaning that the average, faded interference power is 3 dB less than the average, unfaded, signal power, BE = 0.22 as for the AWGN channel. Of course, BE grows linearly with b.

For the case just treated we can do a more exact analysis, as was done by Kavehrad in Ref. 1 (see Case 1). Let us assume that the result of the selection diversity process is deterministic, not random, whereas all the rejected path gains are Rayleigh faded. The method of moments can be applied to get the exact solution for the error probability in

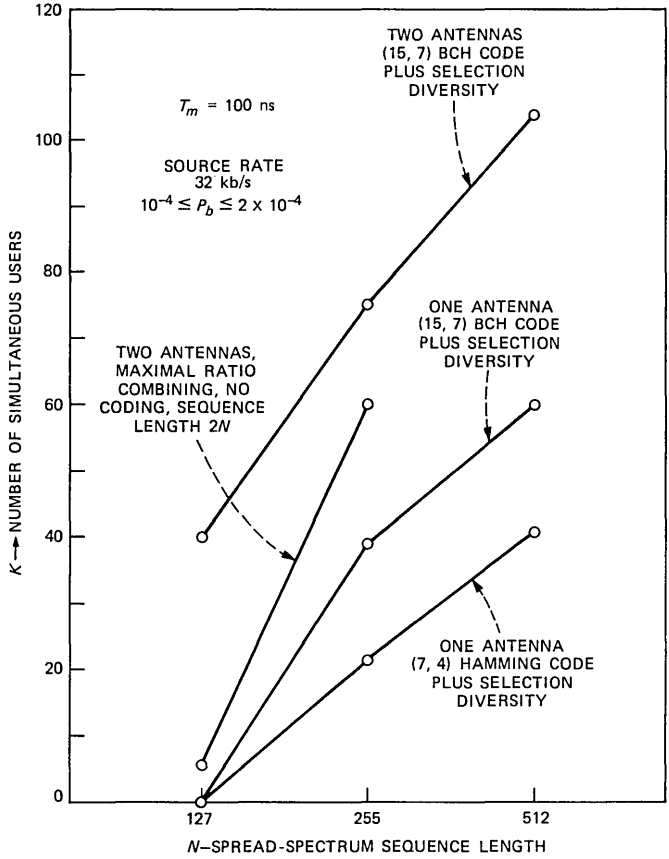


Fig. 12—Number of simultaneous users versus spread-spectrum sequence length when $T_m = 100$ ns.

this case. The theory is similar to that given by Kavehrad[1] and will not be repeated here.

We have summarized the results of all our computations in Fig. 12. The results are all for the Rayleigh faded, discrete multipath model. Note that for a single antenna, the double-error-correction system will support around 40 simultaneous users with a spread-spectrum code length of 255.

7.2 T_m = 250 nanoseconds

We now consider the larger multipath delay spread reported in Ref. 25, which was characteristic of the Holmdel building at AT&T Bell Laboratories. Now $T_m = 250$ ns, a Root Mean Square (RMS) value, a figure that should be used in a larger building. Now for $N = 255$, $R_0 = 32$ kb/s and one gets $L = 4$ in eq. (3) for the discrete multipath model. Other values of L for $T_m = 250$ ns are given in Table IIIc. Note that

Table VIa—The number of simultaneous users in terms of sequence length, N, and the number of base station antennas, L_s. The number of users is given in columns 3, 4, and 5. The order of diversity is $L_s \cdot L$, $T_m = 250$ ns, $P_b \approx 10^{-4}$.

N	L_s	(7, 4) Code	(15, 7) Code	MRC at $2N$	BE (15, 7) Code	BE MRC at $2N$
127	1	10	20	9	0.07	0.04
127	2	30	40	44	0.15	0.17
255	1	26	36	30	0.07	0.06
255	2	48	60	120	0.11	—
512	1	40	50	—	0.05	—

Table VIb—A comparison of the estimate of the number of simultaneous users with $T_m = 250$ ns

N	L_s	Code	L_{max}	K for L_{max}	Average K
127	2	(7, 4)	2	30	27
255	2	(7, 4)	4	48	49
127	2	(15, 7)	3	40	42
255	2	(15, 7)	5	60	68

$N = 512$ would give $L = 8$ or 9 for the coded system, and thus a diversity order that is too large. In this sense our spread-spectrum system has an optimum sequence length or bandwidth. Results similar to those in Table IV are given in Table VIa and are also plotted in Fig. 13. Note that the (15, 7) coded system can support just under 38 simultaneous users with the same antenna diversity and code length (that is, $L_s = 1$ and $N = 255$) as used when $T_m = 100$ ns (see Table IV). However, the multipath diversity is now of order 5 (see Table IIIc). In any case, the number of simultaneous users is about the same. In a simple Time-Division Multiple Access (TDMA) system, one-half the number of users would be lost as the maximum multipath delay spread has been increased by approximately a factor of two. As such, a SSMA system is less sensitive to a change in maximum multipath delay spread than a simple TDMA system would be. However, if more than 40 simultaneous users are needed, the SSMA system also loses. For instance, compare the performance of the double-error-correcting system at $L_s = 2$ when $N = 512$ for $T_m = 100$ ns (Table IV) and when $N = 255$ for $T_m = 250$ ns (Table VIb); the loss is from 108 to 60 simultaneous users.

As in the case when $T_m = 100$ ns we include estimates of K for the random model for L in Table VIa. The trend in the results is essentially the same as was observed in Tables Va and b.

7.3 Multipath outage performance estimate

Up to now we have used the average error probability as the

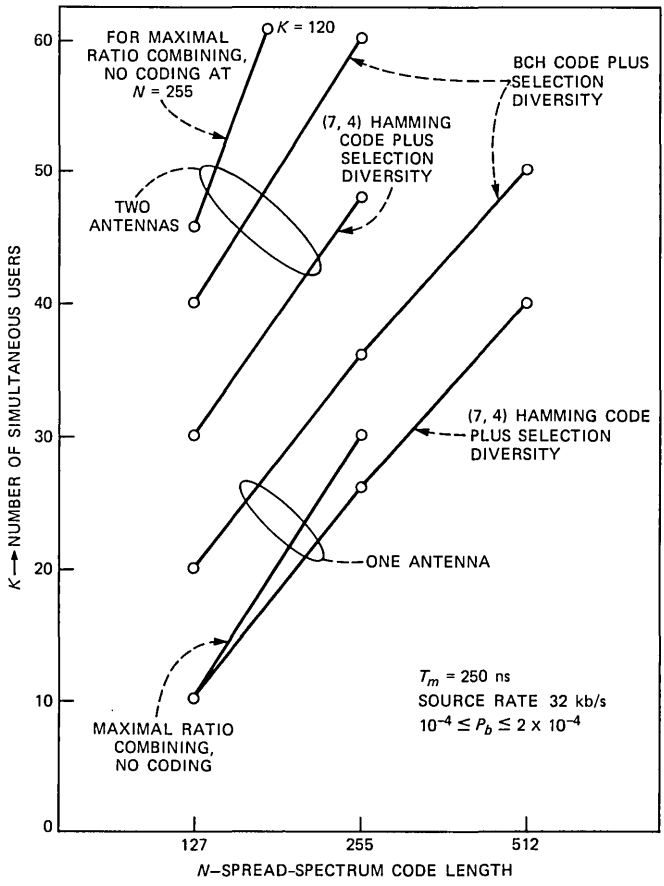


Fig. 13—Number of simultaneous users versus spread-spectrum sequence length when $T_m = 250$ ns.

performance measure. Another measure is to find the distribution of the error probability. This has been used in other radio studies, for example, in Refs. 26 and 27. In the latter study if the probability that the average error probability exceeds the value X is, say, 0.10, the multipath outage is said to be 10 percent. In keeping with our earlier work we take $X = 10^{-4}$.

We will determine the multipath outage for the reference user path gain, β_1, with all other path gains having the Rayleigh statistics as assumed earlier. The computation of the multipath outage follows closely Case 1 of Ref. 1. First β_1 is taken to be fixed, and the error probability is found by averaging the right-hand side of eq. (11) with respect to the sum of the self-plus multiuser interference, using the method of moments. Let us call the result of this calculation $\overline{P_e}(\beta_1)$,

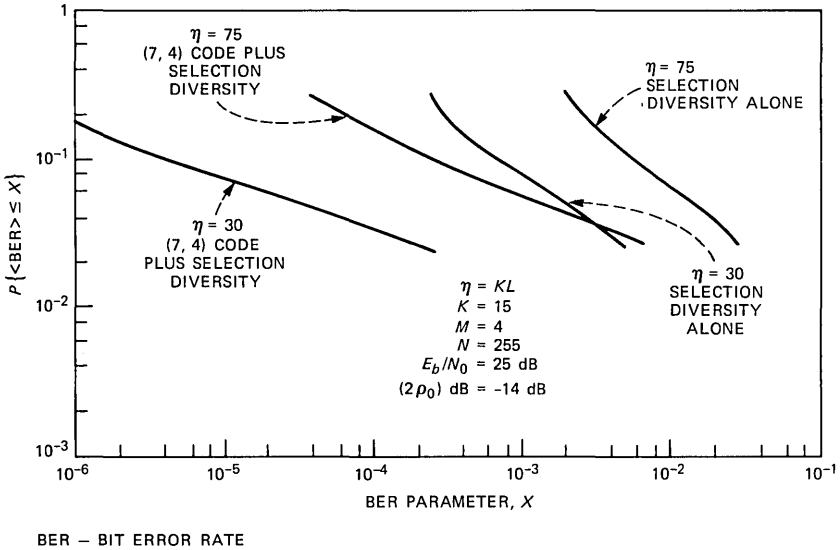
since it depends on the path gain β_1. We then vary β_1 from zero to β_0, where $\overline{P_e}(\beta_0) = X$, the bit error rate parameter. Then

$$P\{\overline{P_e}(\beta_1) \geq X\} = P\{0 \leq \beta_1^2 \leq \beta_0^2\}.$$

The probability $P(0 \leq \beta_1^2 \leq \beta_0^2)$ is easily obtained by integrating the pdf for selection diversity as given in eq. (17).

If coding is involved, $\overline{P_e}(\beta_1)$ changes. For instance, for $X \leq 10^{-2}$, $\overline{P_e}(\beta_1)$ is well approximated by $9\overline{P_e^2}(\beta_1)$, say, for the (7, 4) Hamming code. To find β_0^2 one solves the equation $X = 9\overline{P_e^2}(\beta_0)$ for a given X. Of course, for the same X, β_0 for the coded case is smaller than β_0 for diversity alone, which leads to a lower outage probability.

The results of our computations for the manner just described are shown in Fig. 14. Note that coding is effective in reducing the outage probability for increasing $\eta = KL$. We found that, to a good approximation, the outage probability was only a function of the product, KL. This allows us to estimate the number of users for a fixed multipath outage as we did earlier for the average error probability. The results are shown in Table VIIa for a 10-percent multipath outage and T_m = 100 ns. Note that the results are given for E_b/N_0 = 25 dB. We found that for up to 15 moments, outage probabilities for larger E_b/N_0's were not reliable in these computations. No such problem occurred in computing the average error probability. The number of simultaneous users is about 10 percent less than it would be if the average error probability for E_b/N_0 = 25 dB were used as a performance measure.



BER – BIT ERROR RATE

Fig. 14—Distribution of error probability.

Table VIIa—The number of simultaneous users for a multipath outage probability of 10 percent when $T_m = 100$ ns, $E_b/N_0 = 25$ dB, $2\rho_0 = -14$ dB, and the error probability is parameter $X \approx 10^{-4}$

N	L_s	K (7, 4) Code	K (15, 7) Code	MRC at $2N$
127	2	5	15	0
127	3	20	36	24
255	1	5	15	0
255	2	38	53	53
255	3	45	68	112

Table VIIb—The reduction in outage probability as L_s, the number of antennas, is increased. This reduction is relative to $L_s = 2$, where $K = 38$ for the (7, 4) code and $K = 53$ for both the (15, 7) code and the MRC system.

N	L_s	System	P_o
255	3	Coded	3.5×10^{-2}
255	4	Coded	1.1×10^{-2}
512	3	MRC	10^{-2}
512	4	MRC	4.7×10^{-4}

We note that the Gaussian assumption had the same accuracy as we stated earlier for the average error probability. Accordingly, the MRC results were computed by invoking a Gaussian assumption on the interference. Table VIIb shows results on how the multipath outage can be reduced by increasing antenna diversity. For a large order of diversity, MRC is stronger than selection diversity plus coding in this task.

Our work to date has assumed an average power control so that all user signals arrive at the central station with the same average power. We now let the static attenuation of the reference user be higher than each member for the whole multiuser population. The results for a 10-percent multipath outage are shown in Table VIII. Note that the loss in user population in percent is about the same as the static power loss of the reference user. Thus, SSMA is not sensitive to small deviations in static power control.

VIII. CONCLUSION

Our main conclusion is that direct-sequence, spread-spectrum modulation can give a quite respectable number of simultaneous users for

Table VIII—The percentage reduction in the number of simultaneous users in Table VII as the reference user is subjected to increased attenuation. This is for $T_m = 100$ ns, $X \approx 10^{-4}$, and the multipath outage probability of 10 percent.

$\delta\gamma_0$ dB	$\delta\gamma_0$ Percent	δK (7, 4) Code Percent	δK (15, 7) Code Percent
−0.5	12	21	15
−1.0	26	36	28
−2.0	58	58	47

communication over fading, multipath channels. This is for communication from transportable stations to a base station in a star network using average power control that depends only on the power law exponent and static shadow fading. The inherent diversity of spread-spectrum modulation can be combined with antenna diversity using the simple selection diversity rule to give efficiencies that are comparable to those attained with maximal ratio combining. Fairly high orders of diversity are needed for the latter to be better than selection diversity used with channel coding.

We also have found that a Gaussian assumption regarding the multiuser interference can lead to a maximum error of 20 percent in predicting the number of simultaneous users for a noise floor average error probability of 10^{-4}. This is for a system using selection diversity.

We conclude that spread-spectrum modulation can be less sensitive to a change in maximum multipath delay spread than, say, time-division multiple access would be. Thus the same spread-spectrum modem could possibly be used in either large or small buildings for indoor radio communication.

The main assumptions used in the paper are that demodulation errors are independent and that the multipath model is discrete. The former can be overcome with interleaving. However, the latter must be verified experimentally for indoor, wireless, local area network application.

We have also considered multipath outage as a performance criterion. For an outage of 10 percent we find that the number of simultaneous users is reduced by approximately 10 percent over that predicted by using average error probability as a performance measure.

Although our analysis was for selection diversity or maximal ratio combining and coherent binary phase-shift keying, in practice we would suggest equal gain combining and differential binary phase-shift-keying modulation. We make this suggestion as usually equal gain combining falls in performance between that for selection diversity and maximal ratio combining. Also, differential phase-shift keying

is less troublesome to demodulate in a fading environment such as occurs in IWC.

IX. ACKNOWLEDGMENT

Discussions with David Goodman and L. J. Greenstein have been helpful in our work. Also, Carl-Erik Sundberg is acknowledged for contributing some of the ideas used in Appendix A.

REFERENCES

1. M. Kavehrad, "Performance of Nondiversity Receivers for Spread Spectrum in Indoor Wireless Communications," AT&T Tech. J., *64*, No. 6, Part 1 (July–August 1985), pp. 1181–210.
2. G. H. Golub and J. H. Welsh, "Calculations of Gauss Quadrature Rules," Math. Comput. J., *26* (April 1969), pp. 221–30.
3. Wm. C. Jakes, Jr., *Microwave Mobile Communications*, New York: Wiley, 1974.
4. L. B. Milstein, R. L. Pickholtz, and D. L. Schilling, "Optimization of the Processing Gain of an FSK-FH System," IEEE Trans. Commun., *COM-28* (July 1980), pp. 1062–79.
5. M. K. Simon et al., *Spread Spectrum Communications, 1*, Rockville, Maryland: Computer Science Press, 1984.
6. A. Livine, "Design Considerations for Code Division Multiple Access in Voice/Data Radio Network," Proc. 1984 Military Comm. Conf., October 21–24, Los Angeles, California, pp. 37.2.1–.6.
7. G. Turin, "The Effects of Multipath and Fading on the Performance of Direct-Sequence CDMA Systems," IEEE J. Selected Topics Commun., *SAC-2* (August 1984), pp. 597–603.
8. P. Freret et al., "Applications of Spread-Spectrum Radio to Wireless Terminal Communications," Proc. NTC'80 (June 1980), pp. 69.7.1–.4.
9. S. Nanayakkara and J. B. Anderson, "High Speed Receiver Designs Based on Surface Acoustic Wave Devices," Satellite Commun., *2*, No. 2 (April 1984), pp. 121–8.
10. J. G. Proakis, *Digital Communications*, New York: McGraw-Hill, 1983.
11. P. S. Henry and B. S. Glance, "A New Approach to High Capacity Digital Mobile Radio," B.S.T.J., *60*, No. 8 (October 1981), pp. 1891–904.
12. G. Turin, "Introduction to Spread-Spectrum Anti-Multipath Techniques and Their Application to Urban Digital Radio," Proc. IEEE, *68* (March 1980), pp. 328–53.
13. M. B. Pursley, "Spread-Spectrum Multiple-Access Communication," CISM Course and Lectures No. 265, G. Longo, ed., New York: Springer-Verlag, 1981.
14. J. S. Lehnert and M. B. Pursley, "Multipath Diversity Reception of Coherent Direct-Sequence Spread-Spectrum Communications," Proc. 1983 Conf. Inform. Sci. Syst., The Johns Hopkins University, March 23–25, 1983.
15. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, Second Edition, New York: McGraw-Hill, 1984.
16. M. B. Pursley, "Performance Evaluation for Phase Coded Spread-Spectrum Multiple Access Communication—Part II: Code Sequence Analyses," IEEE Trans. Commun., *COM-25* (August 1977), pp. 800–3.
17. C. E. Sundberg, "Error Probability of Partial Response Continuous-Phase Modulation With Coherent MSK-Type Receiver, Diversity, and Slow Rayleigh Fading in Gaussian Noise," B.S.T.J., *61*, No. 8 (October 1982), pp. 1933–63.
18. V. Pless, *Introduction to the Theory of Error-Correcting Codes*, New York: Wiley-InterScience, 1982.
19. S. Lin and D. J. Costello, Jr., *Error Control Coding: Fundamentals and Applications*, Englewood Cliffs, N.J.: Prentice-Hall, 1983.
20. H. F. A. Roefs and M. B. Pursley, "Correlation Parameters of Random Sequences and Maximal Length Sequences for Spread-Spectrum Multiple-Access Communications," IEEE Trans. Commun., *COM-27* (October 1979), pp. 1597–604.
21. A. A. M. Saleh and R. Valenzuela, private communication.
22. R. Valenzuela, private communication.
23. W. C. Wong, O. G. Jaffee, and D. J. Goodman, private communication.
24. Further Notice of Inquiry and Notice of Proposed Rulemaking, "Authorization of Spread-Spectrum and Other Wideband Emissions Not Presently Provided for in the FCC Rules and Regulations," May 21, 1984.

25. D. M. J. Devasirvatham, "Time Delay Spread Measurements of Wideband Radio Signals Within a Building," Electron. Lett., *20*, No. 23 (November 1984), pp. 950–1.
26. G. J. Foschini and J. Salz, "Digital Communications Over Fading Radio Channels," B.S.T.J., *62*, No. 2 (February 1983), pp. 429–56.
27. B. Glance and L. J. Greenstein, "Frequency-Selective Fading Effects in Digital Mobile Radio With Diversity Combining," IEEE Trans. Commun., *COM-31* (September 1983), pp. 1085–94.
28. G. C. Clark, Jr., and J. B. Cain, *Error-Correction Coding for Digital Communications*, New York: Plenum Press, 1981.

APPENDIX A

Channel Coding Formulations

In this Appendix we derive the formula for the bit error probability used in the paper. Independent channel errors are assumed and only single-error-correcting, (7, 4) Hamming code and double-error-correcting, (15, 7) BCH codes are considered.

A.1 The (7, 4) Hamming code

We first determine the bit error probability for a small channel error probability $p_e = 1 - q_e$. The first term in the power series in p_e for P_{b1} will be denoted as P_{b1}^1. A well-known approximation [see eq. (1-27) of Ref. 28] gives $P_{b1}^1 = dP_s/n$, where $P_s = \binom{7}{2} p_e^2 q_e^5$. Here P_s is the probability, for small p_e, that a code vector is in error and d is the minimum Hamming distance. As $n = 7$ and $d = 3$ we have $P_{b1}^1 = 9p_e^2q_e^5$ for small p_e for the first term in eq. (30).

The weight of a code word is the number of its nonzero symbols. Let A_i be the number of code words of weight i. For the (7, 4) Hamming code $A_0 = A_7 = 1$ and $A_3 = A_4 = 7$ with all other $A_i = 0$. Since the (7, 4) Hamming code is a perfect code, two channel errors always produce a weight-3 code word where we assume the all-zero code word is transmitted. As the code is linear we have no loss in generality. In the weight-3 code words, three code words contain one bit error, three contain two bit errors, and one contains three bit errors, where a bit error refers only to erroneous information bits. Thus, as there are four information bits,

$$P_{b1}^1 = \frac{\binom{7}{2}}{7} p_e^2 q_e^5 \left\{\frac{3}{4} + \frac{6}{4} + \frac{3}{4}\right\},$$

which gives $9p_e^2q_e^5$, as before.

We now find a correction term to P_{b1}^1, which we call P_{b1}^2. When three channel errors result, either a weight-3 or a weight-4 code word is decoded. Now, P_{b1}^2 is the sum of the probabilities of these two cases. The weight-4 code words can be chosen in 28 ways, since each of the seven weight-4 code words has four correctable error patterns. Thus,

$$P_{b1}^{2}(\text{weight-4}) = \frac{28}{7} p_e^3 q_e^4 \left\{\frac{9}{4} + \frac{6}{4} + \frac{1}{4}\right\} = 16 p_e^3 q_e^4,$$

as among the weight-4 code words three have three bit errors, three have two bit errors, and one code word contains a single bit error.

To consider the second component of P_{b1}^{2}, we note that a weight-3 code word is chosen when the channel errors combine with the transmitted all-zero code word to produce a weight-3 code word. Thus,

$$P_{b1}^{2}(\text{weight-3}) = p_e^3 q_e^4 \left\{\frac{3}{4} + \frac{6}{4} + \frac{3}{4}\right\}$$

to give $3p_e^3 q_e^4$. Adding the results for P_{b1}^{2}, we get the correction term in eq. (30) of the paper. We get exactly the same result by applying the approximation dP_s/n to the weight-3 and weight-4 error events. That is, for the weight-3 bit error probability we have $7 \times 3p_e^3 q_e^4/7$ and for weight-4, $28 \times 4p_e^3 q_e^4/7$ to get $P_{b1}^{2} = 19p_e^3 q_e^4$.

A.2 The (15, 7) BCH code

We generated the (15, 7) BCH code with generator polynomial

$$g(x) = 1 + x^4 + x^6 + x^7 + x^8.$$

This gave $A_0 = 1$, $A_5 = 18$, and $A_6 = 28$, which are the only components of the weight distribution we will need. Now A_5 has five code words with one bit error, five with two, six with three, and two with four. Thus,

$$P_{b2}^{1}(\text{weight-5}) = \frac{455}{18} p_e^3 q_e^{12} \left\{\frac{5}{7} + \frac{10}{7} + \frac{18}{7} + \frac{8}{7}\right\} = \frac{2470}{18} p_e^3 q_e^{12},$$

since there are $\binom{15}{3} = 455$ error patterns with three channel errors. Also,

$$P_{b2}^{1}(\text{weight-6}) = \frac{455}{28} p_e^3 q_e^{12} \left\{\frac{2}{7} + \frac{16}{7} + \frac{39}{7} + \frac{16}{7} + \frac{5}{7}\right\} = \frac{5070}{28} p_e^3 q_e^{12},$$

since in A_6 two code words have one bit error, eight have two, thirteen have three, four have four, and one code word has one bit error.

To combine our two values of P_{b2}^{1}, we use the density of the code words. For each of the 128 code words there are $\binom{15}{2} = 105$ correctable double-error patterns and 15 correctable single-error patterns. Thus the number of channel outputs within a distance two of code words is

121 × 128. There are 2^{15} channel outputs in all, to give a probability of 0.47 of falling within a decoding sphere of radius two. We assume 47 percent of the space around the zero code word is filled with weight-5 code words. In the remaining 53 percent we assume weight-5 and weight-6 code words are chosen with equal probability. Thus,

$$P_{b2}^1 = P_{b2}^1(\text{weight-5}) \times 0.735 + P_{b2}^1(\text{weight-6}) \times 0.265 = 150 p_e^3 q_e^{12},$$

which agrees well with $P_{b2}^1 = P_s d/15$, as now $d = 5$ and $P_s = \binom{15}{3} p_e^3 q_e^{12}$.

To get the correction term we assume that the four channel errors produce either a weight-5 code word or a weight-6 code word. The weight-5 codes have five correctable error patterns of weight-4 per code word and thus $18 \times 5 = 90$ weight-4 channel outputs in their decoding spheres. For weight-6 code words we have $\binom{6}{2} = 15$ correctable error patterns and thus $28 \times 15 = 420$ channel outputs in weight-6 decoding spheres. This leaves $\binom{15}{4} - 510 = 855$ error patterns and we assume they are equally divided between weight-5 and weight-6 code words. Averaging over the bit errors in weight-5 and weight-6 codes gives

$$P_{b2}^2 = \left(\frac{518 \times 41}{18 \times 7} + \frac{848 \times 78}{28 \times 7} \right) p_e^4 q_e^{11} = 506 p_e^4 q_e^{11},$$

where the first term is for weight-5 code words. Use of the dP_s/n approximation gives

$$P_{b2}^2 = \left(\frac{518 \times 5}{15} + \frac{848 \times 6}{18} \right) p_e^4 q_e^{11} = 512 p_e^4 q_e^{11},$$

which is in good agreement with the result just given above.

AUTHORS

Mohsen Kavehrad, B.S. (Electrical Engineering), 1973, Tehran Polytechnic Institute; M.S. (Electrical Engineering), 1975, Worcester Polytechnic Institute; Ph.D. (Electrical Engineering), 1977, Polytechnic Institute of New York; Fairchild Industries, 1977–1978; GTE, 1978–1981; on the faculty of Northeastern University, 1981–1984; AT&T Bell Laboratories, 1981—. At AT&T Bell Laboratories Mr. Kavehrad is a member of Communications Methods Research Department. His research interests are digital communications and computer networks. He is a Technical Editor for the IEEE Communications Magazine. He established and was the Chairman of the IEEE Communications Chapter of New Hampshire in 1984. Member, IEEE, Sigma Xi.

Peter J. McLane, B.A.Sc., 1965, University of British Columbia; M.S.E.E., 1966, University of Pennsylvania; Ph.D., 1969, University of Toronto; Queen's University, 1969—. Mr. McLane joined the National Research Council of Canada in 1967, and in 1969 joined Queen's University, where he is currently a Professor of Electrical Engineering. Since 1984 he has been visiting the Communications Methods Research Department at AT&T Bell Laboratories. His research interests are in the area of Communication Theory and its application. He is a former Associate Editor for the IEEE Communications Magazine and is currently the Associate Editor for Communication Theory, IEEE Transactions on Communications. He is also the Guest Associate Editor for the IEEE Journal on Special Topics in Communications (Topic, VLSI in Communications). Member, IEEE, Eta Kappa Nu.

AT&T Technical Journal
Vol. 64, No. 8, October 1985
Printed in U.S.A.

Nonlinear Input-Output Maps and Approximate Representations*

By I. W. SANDBERG†

(Manuscript received May 20, 1985)

An approximation theorem is given for causal time-invariant nonlinear maps that take one set of functions defined on [0, ∞) into another. The theorem is used to show that, under some typically very reasonable conditions, an input-output map can be approximated arbitrarily well in a meaningful sense by a finite Volterra series, even though it may not have a Volterra series expansion. The set of inputs on which the approximation holds need not be compact, and the inputs need not be continuous.

I. INTRODUCTION

In this paper an approximation theorem is given for causal time-invariant nonlinear maps that take one set of functions defined on [0, ∞) into another. The theorem is used to show that, under some typically very reasonable conditions, an input-output map can be approximated arbitrarily well in a meaningful sense by a finite Volterra series, even though it may not have a Volterra series expansion. The set of inputs on which the approximation holds need not be compact. A more detailed introduction follows.

1.1 Background

Researchers have long been interested in a variety of questions concerning the mathematical representation of systems that need not

* This paper was presented at the Midwest Symposium on Circuits and Systems, Louisville, Kentucky, August 19–20, 1985.

† AT&T Bell Laboratories.

be linear. In recent years much has been learned about the existence, determination, and properties of power-series-like expansions for expressing a system's outputs in terms of its inputs (see, for instance, Refs. 1 through 7). An important example of the form of such an expansion is

$$w(t) = \sum_{q=1}^{\infty} \int_0^t \cdots \int_0^t k_q(\tau_1, \cdots, \tau_q) u(t - \tau_1) \cdots u(t - \tau_q) d\tau_1 \cdots d\tau_q, \qquad t \geq 0, \quad (1)$$

in which w is the output, u is the input, the kernels k_q are functions determined by the system, and u is drawn from a set of bounded real-valued inputs such that the right side of (1) converges uniformly in t for every input. While certain more general versions of (1), where the k_q are *symbolic* functions and vector-valued inputs are taken into account, are frequently needed, in all cases one has

$$w = \sum_{q=1}^{\infty} K_q(u), \qquad u \in U, \quad (2)$$

in which U is a set of inputs and each K_q is a homogeneous map of degree q. Under weak assumptions these K_q are *uniquely* determined, and as such are very special associates of the system represented by (2).

The right side of (1) is an example of what is often called a Volterra series. Actually, Volterra considered not (1) but related expansions in which the integration limits are constants and the $u(t - \tau_j)$ are replaced with $u(\tau_j)$. These expansions were used by Volterra as a model of a nonlinear functional in his path-breaking studies of operations on functionals. A comprehensive account of this work is given in Ref. 8, where attention is directed to a representation result (Ref. 8, page 20) due to Fréchet that concerns, in particular, the uniform approximation of continuous functionals on compact sets, using a finite number of terms in Volterra's series. Volterra also mentions the analogy between this aspect of Fréchet's result and the Weierstrass approximation theorem for continuous real functions on a real compact interval.

While Fréchet's Weierstrass-like result is certainly interesting and important, it does have significant limitations with regard to the representation of input-output maps: (1) it concerns *approximations* rather than expansions in the usual sense, (2) these approximations are on *compact* sets, (3) it directly concerns functionals rather than mappings from one function space to another, etc. These limitations, as well as those of the more general representation result of Fréchet described in Ref. 8, p. 20, do not appear to have been always appreci-

ated by early writers concerned with applications of Volterra series, who sometimes cited Ref. 8 as though it contained a justification for the use of relations of the form (1).

There can be very basic differences between an arbitrarily good approximation and an expansion, and, more to the point, between knowing that one, as opposed to the other, exists. For example, an approximation known to exist may not have properties that facilitate its determination. Nevertheless, existence results concerning approximations can sometimes be useful, especially when an expansion does not exist. Thus, it is clearly of interest to consider the extent to which input-output maps of systems can be approximated in some meaningful sense by a finite sum,

$$\sum_{q=1}^{Q} \int_0^t \cdots \int_0^t k_q(\tau_1, \cdots, \tau_q) u(t-\tau_1) \cdots u(t-\tau_q) d\tau_1 \cdots d\tau_q, \qquad t \geqslant 0, \quad (3)$$

of terms of the form that appear on the right side of (1), or by a finite sum of suitably more general terms if u is vector valued. Of course, approximations involving larger classes of finite sums of iterated integrals can be of interest too.

Related questions have in fact been considered for many years,[9-15] and as one might expect, the main mathematical tool that has been used is the Stone-Weierstrass theorem. In this earlier work the input signals considered are assumed to belong to a Hilbert space (e.g., an L_2 space) and/or to be defined on only a *finite* interval, and the inputs are taken to belong to a *compact* set. In contrast, in Ref. 16 an approximation result is given for input-output maps that act between certain subsets of the Banach space $C(\mathbb{R})$ of bounded, continuous, real-valued functions defined on the doubly infinite interval $(-\infty, \infty)$, with the usual norm. The maps considered there are assumed to be time invariant and to have a "fading-memory" property that enables one to prove that a certain set of functions defined on $(-\infty, 0]$ is compact. The extent to which the results in Ref. 16 bear on the main problem considered in this paper, where the input and output signals are defined on $[0, \infty)$, is not discussed in Ref. 16. Although the results in this paper are considerably different from those in Ref. 16, there are some similarities: an approximately-finite memory hypothesis (related to hypotheses in Ref. 17, Section 2.2) plays a central role, and we too depend on a form of the Stone-Weierstrass theorem. On the other hand, the compact sets with which we deal are always sets of functions defined on a *finite* interval $[0, \omega]$.

1.2 Outline of this paper's results

In Section II, attention is focussed on a class of causal time-invariant

maps G that take S into S_0, where S and S_0 are sets of signals (i.e., sets of functions) defined on $[0, \infty)$, and the elements of S_0 are real valued. The maps G are assumed to possess a factorization FH, where H takes S into S_1, and F maps S_1 into S_0, where S_1 is a third set of signals on $[0, \infty)$. Certain hypotheses on S, S_0, S_1, and the factors F and H are introduced in Theorem 1 of Section 2.2. Under those conditions, the theorem shows that given any $\epsilon > 0$, there are a constant $\Delta \geq 0$, and a map P having an important special form (that involves a real polynomial p in several variables together with a certain "fundamental set " of maps) such that the approximation

$$|G(u)(t) - (PH)(u_{\max[0,t-\Delta]})(t)| < \epsilon, \qquad t \geq 0$$

holds for all $u \in S$, where u_ω for arbitrary nonnegative ω is defined by $u_\omega(t) = 0$ if $0 \leq t < \omega$, and $u_\omega(t) = u(t)$ for $t \geq \omega$. One of the main hypotheses used is that the memory of G is "approximately finite," to the extent that for any $\delta_0 > 0$ there is a $\delta > 0$ for which

$$|Gu(t) - Gu_{\max[0,t-\delta]}(t)| < \delta_0$$

for $t \geq 0$ and $u \in S$. This hypothesis can be shown to be satisfied in many cases of interest. More will be said about this later.

The hypotheses of Theorem 1 are of an abstract nature and so is its conclusion. The theorem is used as a "tool theorem" in the proof of Theorem 2 in Section 2.4 which addresses a case that is of direct interest in applications. In Theorem 2, the memory of G is assumed to be approximately finite, S is taken to be a set of uniformly bounded vector-valued functions on $[0, \infty)$, S_1 is a similar set, H is a convolution, and F is causal, time invariant, and continuous in a certain typically reasonable sense (see the theorem for additional details). Assume now for the sake of simplifying the discussion that the elements of S are *scalar* valued. According to the theorem, under the conditions stated there, Gu can be uniformly approximated arbitrarily well on S by a finite sum of the form (3). Again, the reader is referred to the theorem for the details. Related results concerning discrete-time cases and composites of maps that have approximately-finite memory are given in Sections 2.5 and 2.3, respectively.

The case considered in Theorem 2 arises often. This is discussed in Appendix B, where an important class of input-output maps is addressed, and where a technique for showing that the memory of a nonlinear map is approximately finite is illustrated.

II. INPUT-OUTPUT MAPS AND APPROXIMATIONS

2.1 Preliminaries

Throughout Section II, V is a linear space, Ω denotes the interval $[0, \infty)$, t and ω are elements of Ω, S and S_0 are two sets of functions

on Ω, the elements of S and S_0 take values in V and $(-\infty, \infty)$, respectively, and G is a map from S to S_0.

We use S_1 to denote a third set of functions on Ω; these take values in a normed linear space V_1. It is assumed that

$$G = FH,$$

where H maps S into S_1 and F takes S_1 into S_0.

The set S, which is our set of inputs, is assumed to have the following properties:

(i) $u \in S \Rightarrow (u)_\omega \in S$ for each ω, where $(u)_\omega(t) = u(t)$ for $t \leq \omega$, and $(u)_\omega(t) = 0$ (here the zero element of V) otherwise.

(ii) $u \in S \Rightarrow (T_\omega u) \in S$ for $\omega \neq 0$, in which T_ω is defined by $(T_\omega u)(t) = 0$ $(0 \leqslant t < \omega)$ and $(T_\omega u)(t) = u(t - \omega)$ for $t \geqslant \omega$.

(iii) $u \in S \Rightarrow u_\omega \in S$ for $\omega \neq 0$, where u_ω is given by $u_\omega(t) = 0$ for $0 \leqslant t < \omega$ and $u_\omega(t) = u(t)$ for $t \geqslant \omega$. [Note the distinction between u_ω and $(u)_\omega$ of Property (i).]

(iv) $\omega \neq 0$ and $u \in S$ with $u(t) = 0$ $(0 \leqslant t < \omega) \Rightarrow v \in S$, where $v(t) = u(t + \omega)$, $t \geqslant 0$.

Notice that (i)–(iv) simply require that S be closed under certain elementary operations. With regard to S_1 we assume that

(v) Properties (ii) and (iii) hold with S replaced with S_1.

We use the standard definitions of causality and time invariance. That is, a map M from S to S_0, from S to S_1, or from S_1 to S_0 is *causal* if u_1 and u_2 in the domain of M with $u_1(t) = u_2(t) (0 \leqslant t \leqslant \omega)$ always implies that $(Mu_1)(t) = (Mu_2)(t)$ for $0 \leqslant t \leqslant \omega$; M is *time invariant* if $\omega \neq 0$ and u in the domain of $M \Rightarrow (MT_\omega u)(t) = 0$ for $0 \leqslant t < \omega$ and $(MT_\omega u)(t) = (Mu)(t - \omega)$, $t \geqslant \omega$. Also, by $M \in \mathscr{A}(S)$ or $M \in \mathscr{A}(S_1)$ we mean that the domain of M is S or S_1, respectively, and that M has "approximately-finite memory" in the sense that for each constant $\delta_0 > 0$ there is a positive $\delta \in \Omega$ such that

$$|(Mu)(t) - (Mu_{\max[0,t-\delta]})(t)| < \delta_0, \qquad t \geqslant 0$$

for all u in the domain of M. Here $|\cdot|$ denotes simply the absolute value if the range of M is S_0; it denotes the norm in V_1 otherwise.

The set of functions x defined on $[0, \omega]$ with values in V_1 such that $x(t) = y(t)$ $(0 \leqslant t \leqslant \omega)$ for some $y \in S_1$ is denoted by $(S_1)_\omega$ for each ω. We use $H(S)_\omega$ to stand for the set of functions x in $(S_1)_\omega$ such that $x(t) = (Hu)(t)$ $(0 \leqslant t \leqslant \omega)$ for some $u \in S$. It is assumed in Theorem 1 (below) that the $H(S)_\omega$ have the following property:

(vi) There is a family of metric spaces $\{(X_\omega, \rho_\omega) : \omega > 0\}$ such that for each $\omega > 0$ we have $H(S)_\omega \subset X_\omega \subset (S_1)_\omega$ and X_ω is compact (i.e., compact in itself) with respect to the metric ρ_ω.

For $\omega \neq 0$ and each causal map M from S_1 to S_0, M_ω denotes the functional on $(S_1)_\omega$ defined by

$$M_\omega x = (My)(\omega), \qquad x \in (S_1)_\omega, \tag{4}$$

where $y \in S_1$ satisfies $x(t) = y(t)$, $0 \leqslant t \leqslant \omega$. The following is assumed with regard to F in Theorem 1 (below):

(vii) F is causal and for each $\omega \neq 0$, F_ω is *continuous* on X_ω with respect to ρ_ω [where X_ω and ρ_ω are described in (vi)].

Finally, by a *fundamental set* $\{F_\alpha : \alpha \in \Lambda\}$ of maps from S_1 to S_0 relative to (vi), we mean that the F_α are causal and time invariant, and that for $\omega \neq 0$ the corresponding family $\{F_{\alpha\omega} : \alpha \in \Lambda\}$ of functionals on $(S_1)_\omega$ is continuous on X_ω with respect to ρ_ω, and separates the points of X_ω. (By "separates the points of X_ω" is meant Ref. 18, p. 41 that for each pair of distinct elements x_1 and x_2 of X_ω there is an $\alpha \in \Lambda$ such that $F_{\alpha\omega}x_1 \neq F_{\alpha\omega}x_2$.)

2.2 The main approximation result

In this section we prove the following:

Theorem 1: Let (i)–(vii) be met, with F and H causal and time invariant, and with $G \in \mathscr{A}(S)$. *Suppose that there is a fundamental set* $\{F_\alpha : \alpha \in \Lambda\}$ *of maps from* S_1 *to* S_0 *relative to (vi). Then for each* $\epsilon > 0$ *there are a* $\Delta \in \Omega$, *a positive integer k, elements* $F_{\alpha_1}, \cdots, F_{\alpha_k}$ *of* $\{F_\alpha : \alpha \in \Lambda\}$, *and a real polynomial p in k real variables with* $p(0, \cdots, 0) = 0$ *such that*

$$|\, G(u)(t) - (PH)(u_{\max[0,t-\Delta]})(t) \,| < \epsilon, \qquad t \geqslant 0 \tag{5}$$

for every $u \in S$, *where P is the map from* S_1 *into* S_0 *given by* $(Py)(t) = p[F_{\alpha_1}(y)(t), \cdots, F_{\alpha_k}(y)(t)]$, $t \geqslant 0$ *for* $y \in S_1$. *In addition, the map* $Q : S \to S_0$ *defined by* $(PH)(u_{\max[0,t-\Delta]})(t) = (Qu)(t)$ *for* $u \in S$ *and* $t \geqslant 0$ *is causal and time invariant.*

2.2.1 Proof of Theorem 1

Proof: Given ϵ, choose a positive $\Delta \in \Omega$ so that

$$|\, (Gu)(t) - G(u_{\max[0,t-\Delta]})(t) \,| < \epsilon/2, \qquad t \geqslant 0 \tag{6}$$

for $u \in S$. Observe that S contains an element θ such that $\theta(t) = 0$ for $t \geqslant 0$. By the time-invariance of H, $H(S)$ also contains such an element, and therefore there is $e \in X_\Delta$ such that $e(t) = 0$ for $t \in [0, \Delta]$. By the causality and time invariance of F, we have $F_\Delta e = 0$. Using a version of the Stone-Weierstrass theorem (see Ref. 18, p. 46), Condition (vii), and the hypothesis that there is a fundamental set $\{F_\alpha : \alpha \in \Lambda\}$ of maps from S_1 to S_0 relative to (vi),* there are a positive integer k, a polynomial p as described, and elements $F_{\alpha_1}, \cdots, F_{\alpha_k}$ of the fundamental set such that

$$|\, F_\Delta x - P_\Delta x \,| < \epsilon/2, \qquad x \in X_\Delta,$$

* It was necessary to establish that $F_\Delta e = 0$ because $F_{\alpha\Delta}e = 0$ for all $\alpha \in \Lambda$ (see Ref. 18, Theorem 5).

where the functional P_Δ is the associate [see (4)] of P described in the statement of the theorem. Thus, using $H(S)_\Delta \subset X_\Delta$, we have

$$|(FHu)(\Delta) - (PHu)(\Delta)| < \epsilon/2, \qquad u \in S. \tag{7}$$

Now let any $u \in S$ and $t \in \Omega$ be given. Suppose first that $t > \Delta$. Let v be defined by $v(\tau) = u_{(t-\Delta)}[\tau + (t - \Delta)]$, $\tau \geqslant 0$ [notation of Condition (iii)]; here we have used Conditions (iii) and (iv). By the time invariance of $G = FH$ and of PH in (7), one has $G[u_{(t-\Delta)}](t) = G(v)(\Delta)$ and $(PH)[u_{(t-\Delta)}](t) = (PH)(v)(\Delta)$. Thus, by (7),

$$|G[u_{(t-\Delta)}](t) - (PH)[u_{(t-\Delta)}](t)| < \epsilon/2.$$

Using this and (6),

$$|G(u)(t) - (PH)[u_{\max(0,t-\Delta]}](t)| \leqslant |G(u)(t) - G(u_{\max[0,t-\Delta]})(t)|$$

$$+ |G(u_{\max[0,t-\Delta]})(t) - (PH)(u_{\max[0,t-\Delta]})(t)| < \epsilon/2 + \epsilon/2 = \epsilon.$$

Suppose now that $t < \Delta$. Then by Conditions (i) and (ii) and the causality and time invariance of G, we see that $G(u)(t) = G[(u)_t](t) = G[T_{(\Delta-t)}(u)_t](\Delta)$ [notation of Conditions (i) and (ii)], and similarly $(PH)(u)(t) = (PH)[T_{(\Delta-t)}(u)_t](\Delta)$. Thus, using (7),

$$|(Gu)(t) - (PH)(u)(t)| < \epsilon/2. \tag{8}$$

Since (8) holds also for $t = \Delta$, and obviously $u = u_{\max[0,t-\Delta]}$ when $t \leq \Delta$, we have (5). At this point it suffices to prove the following:

Lemma: Let K be a causal time-invariant map of S into S_0, and with $\Delta \in \Omega$, let M be the map from S to S_0 given by

$$(Mu)(t) = K(u_{\max[0,t-\Delta]})(t), \qquad t \geqslant 0$$

for $u \in S$. Then M is causal and time invariant.

Proof: Let u_a and u_b in S satisfy $u_a(\tau) = u_b(\tau)$ for $0 \leqslant \tau \leqslant \omega$, and let $t \in [0, \omega]$. Since $(u_{a\max[0,t-\Delta]})_t = (u_{b\max[0,t-\Delta]})_t$ and K is causal, it is clear that $(Mu_a)(t) = K[(u_{a\max[0,t-\Delta]})_t](t) = K[(u_{b\max[0,t-\Delta]})_t(t) = (Mu_b)(t)$, showing that M is causal.

Now let $u \in S$ and let ω in Ω be nonzero. For $t < \omega$, $(MT_\omega u)(t) = K[\{T_\omega u\}_{\max[0,t-\Delta]}](t) = 0$, because K is time invariant and $\{T_\omega u\}_{\max[0,t-\Delta]}(\tau) = 0$ for $\tau < \omega$. Suppose that $t \geqslant \omega$. Since (as can easily be verified) $\{T_\omega u\}_{\max[0,t-\Delta]} = T_\omega u_{\max[0,t-\omega-\Delta]}$, one has $(MT_\omega u)(t) = K(T_\omega u_{\max[0,t-\omega-\Delta]})(t) = K(u_{\max[0,t-\omega-\Delta]})(t - \omega) = (Mu)(t - \omega)$, by the time invariance of K. Thus M is time invariant. This completes the proof of the lemma and of the theorem.

2.3 Comments

All of the material in Sections 2.1 and 2.2 remains valid if Ω is replaced throughout with $\{0, 1, \cdots\}$ (with the understanding that then $[0, \omega]$ means $\{0, \cdots, \omega\}$).

The following result concerning the hypothesis that $G \in \mathscr{A}(S)$ is frequently useful. By "$F \in \mathrm{Lip}(H(S))$" below is meant that there is a positive constant c such that

$$|(Fy_1)(t) - (Fy_2)(t)| \leqslant c \sup_{\tau \in [0,t]} |y_1(\tau) - y_2(\tau)|$$

for $t \geqslant 0$ and y_1 and y_2 in $H(S)$, where $|\cdot|$ on the right side denotes the norm in V_1.

Proposition: Let (i)–(v) be met,* with $H \in \mathscr{A}(S)$, $F \in \mathscr{A}(S_1)$, and $F \in \mathrm{Lip}[H(S)]$. Then $G \in \mathscr{A}(S)$.

The proposition is proved in Appendix A. For an application, see Appendix B.

2.4 Approximations and finite Volterra series

In the following theorem, n and p are arbitrary positive integers and $L_\infty(n)$ and $L_\infty(p)$, respectively, denote the normed linear spaces of real n-vector valued and real p-vector valued Lebesgue measurable functions u defined on Ω such that $\|u\| \triangleq \sup_t |u(t)| < \infty$, in which $|u(t)| = \max_j |u_j(t)|$ and $u_j(t)$ stands for the jth component of $u(t)$. By a "vector," we mean a column vector. Also, for any positive integer q and any q n-vectors $a_1, \cdots, a_q$, we use $\chi[a_1, \cdots, a_q]$ to denote the vector of order n^q whose elements are the n^q distinct products $(a_1)_{\lambda_1} \cdots (a_q)_{\lambda_q}$, corresponding to distinct sequences $\lambda_1, \cdots, \lambda_q$ with each λ_j drawn from $\{1, \cdots, n\}$, arranged in an arbitrary predetermined order that depends only on q and n. Of course, $\chi[a_1, \cdots, a_q]$ is simply the product $a_1 \cdots a_q$ if $n = 1$. Finally, we use $K(q, s)$ (q, s positive; q an integer) to denote the set of all functions k from $[0, \infty)^q$ to the $1 \times n^q$ matrices such that

$$k_j(\tau_1, \cdots, \tau_q) = \sum_{r=1}^{R_j} \prod_{i=1}^{q} \phi_{jir}(\tau_i), \tag{9}$$

for all $\tau_1, \cdots, \tau_q$ and all $j \in \{1, \cdots, n^q\}$, where $R_j < \infty$ and the ϕ_{jir} are real valued and continuous on $[0, s]$ and vanish on (s, ∞). [Notice that $K(q, s)$ is simply a set of row-matrix-valued functions whose elements have a certain nice finite sum of products representation.]

Theorem 2: Let $G \in \mathscr{A}(S)$, with $S = \{u \in L_\infty(n): \|u\| \leqslant \beta\}$, where β is a positive constant. Assume that H is defined by

$$(Hu)(t) = \int_0^t h(t - \tau)u(\tau)d\tau, \qquad t \geqslant 0$$

for $u \in S$, where h is a real $p \times n$ matrix-valued function on Ω such that each h_{ij} is (Lebesgue) integrable on Ω. Take S_1 to be $\{y \in L_\infty(p): \|y\|$

* For the sake of ease of exposition, (i)–(v) are assumed to be satisfied. However, only (iii) and (iii) with S replaced with S_1 are used.

$\leqslant \beta_1\}$, *in which* β_1 *is any number that satisfies* $\sup_t |(Hu)(t)| \leqslant \beta_1$ *for* $u \in S$.* *Assume also that* F *is causal and time invariant on* S_1, *and that* F *satisfies the continuity condition that given a continuous* y *in* S_1, *and numbers* $t \in (0, \infty)$ *and* $\delta_1 > 0$, *there is a* $\delta_2 > 0$ *such that*

$$|(Fy)(t) - (Fz)(t)| < \delta_1$$

whenever $z \in S_1$, z *is continuous, and* $\max_{\tau \in [0,t]} |y(\tau) - z(\tau)| < \delta_2$. *Under these conditions, given any* $\epsilon > 0$, *there is a positive integer* Q, *an* $s \in (0, \infty)$, *and elements* k_q *of* $K(q, s)$ $(1 \leqslant q \leqslant Q)$ *such that*

$$|(Gu)(t) - (Vu)(t)| < \epsilon, \qquad t \geqslant 0 \tag{10}$$

for all $u \in S$, *where*

$$(Vu)(t) = \sum_{q=1}^{Q} \int_0^t \cdots \int_0^t k_q(\tau_1, \cdots, \tau_q)\chi[u(t - \tau_1), \cdots, u(t - \tau_q)]d\tau_1 \cdots d\tau_q. \tag{11}$$

2.4.1 Proof of Theorem 2

Proof: We use Theorem 1. Conditions (i)–(v) are met with F and H causal and time invariant, and with $G \in \mathscr{A}(S)$.

Consider Condition (vi). Let ω be any positive number. For $x \in H(S)_\omega$,

$$x(t) = \int_0^t h(t - \tau)u(\tau)d\tau, \qquad t \in [0, \infty)$$

for some $u \in S$. In particular,

$$\sup_{t \in [0,\omega]} |x(t)| \leqslant \beta_1, \tag{12}$$

and there is a function λ from $(-\infty, \infty)$ to $[0, \infty)$, which depends only on h, such that $\lambda(\alpha) \to 0$ as $\alpha \to 0$, and

$$|x(t + \alpha) - x(t)| \leqslant \beta\lambda(\alpha) \tag{13}$$

for t and $(t + \alpha)$ in $[0, \omega]$. The existence of such a λ follows directly from the result (see Ref. 19, p. 14) that

$$\int_{-\infty}^{\infty} |r(t + \alpha) - r(t)|dt \to 0 \quad \text{as} \quad \alpha \to 0$$

when r is integrable on $(-\infty, \infty)$. Now let $\{X_\omega, \rho_\omega\}$ be the metric space of all functions x from $[0, \omega]$ to the real p-vectors such that (12) as well as (13) are satisfied, and

* Our conditions here on S and S_1 are clearly consistent with (i)–(v) of Section 2.1.

$$\rho_\omega(x_1, x_2) = \max_{t\in[0,\omega]} |x_1(t) - x_2(t)|.$$

This space is easily seen to be closed. For $p = 1$ its elements are uniformly bounded and equicontinuous. Thus, the space is compact for $p = 1$. Using this fact and, for example, the proposition that compactness in a metric space is equivalent to sequential compactness, it follows that the space is in fact compact for any positive integer p. Clearly, $H(S)_\omega \subset X_\omega \subset (S_1)_\omega$ which shows that (vi) holds.

By the continuity condition on F in Theorem 2, (vii) is satisfied.

Now let $\{F_\alpha : \alpha \in \Lambda\}$ be the set of all maps M defined on S_1 having the representation

$$(My)(t) = \int_0^t m(t - \tau)y(\tau)d\tau, \qquad t \geq 0 \tag{14}$$

where m is a real $(1 \times p)$ matrix-valued function on $[0, \infty)$ such that for any component m_j of m there is a real $\sigma > 0$ for which m_j is continuous on $[0, \sigma]$ and vanishes on (σ, ∞). Let $\omega > 0$, and observe that the $F_{\alpha\omega}$ are continuous on X_ω with respect to ρ_ω.

To see that they separate points of X_ω, let x_1 and x_2 be distinct points of X_ω. Let $i \in \{1, \cdots, p\}$ be such that $x_{1i}(t) - x_{2i}(t) \not\equiv 0$ on some subinterval of $[0, \omega]$. Let $\alpha \in \Lambda$ be such that $(F_\omega y)(t)$ is given by the right side of (14) with $m_i(t) = [x_{1i}(\omega - t) - x_{2i}(\omega - t)]$ for $t \in [0, \omega]$ and $m_i(t) = 0$ otherwise, and with m_j vanishing on $[0, \infty)$ for $j \neq i$. Then

$$F_{\alpha\omega}x_1 - F_{\alpha\omega}x_2 = \int_0^\omega [x_{1i}(\tau) - x_{2i}(\tau)]^2 d\tau > 0.$$

Thus $\{F_\alpha : \alpha \in \Lambda\}$ is a fundamental set in the sense of Theorem 1, and by Theorem 1 given $\epsilon > 0$ there are Δ, k, $F_{\alpha_1}, \cdots, F_{\alpha_k}$, p, and P as described there such that (10) holds with $(Vu)(t) = (PH)(u_{\max[0,t-\Delta]})(t)$.

For $t \geq 0$, we have

$$(PH)(u_{\max[0,t-\Delta]})(t)$$
$$= p[(F_{\alpha_1}H)u_{\max[0,t-\Delta]}(t), \cdots, (F_{\alpha_k}H)u_{\max[0,t-\Delta]}(t)]. \tag{15}$$

It is not difficult to verify that for any $j \in \{1, \cdots, k\}$ the operator $(F_{\alpha_j}H)$ is equivalent to a convolution C whose $1 \times n$ matrix-valued kernel c has elements that are continuous and integrable on Ω.

Also, one finds that

$$\int_0^t c(t - \tau)u_{\max[0,t-\Delta]}(\tau)d\tau = \int_0^t b(t - \tau)u(\tau)d\tau, \qquad t \geq 0,$$

where $b(t) = c(t)$ for $t \in [0, \Delta]$ and $b(t)$ equals the $1 \times n$ zero matrix

otherwise. This together with (15), and just the observation that products of integrals can be written as iterated integrals, shows that V is as described in Theorem 2.

2.5 Comments

Cases in which the conditions of Theorem 2 are met arise often in applications. This is illustrated in Appendix B where the theorem is used to show that an important large class of input-output maps have finite Volterra series approximations.

In the proof of Theorem 2, the F_α are taken to be linear operators. It is clear that related additional approximation theorems can be obtained by allowing the F_α to be nonlinear.

Equation (15) in the proof of Theorem 2 shows that G in the theorem can be approximated arbitrarily well by a linear dynamic subsystem followed by a memoryless nonlinear subsystem with "polynomial nonlinearities." The existence of approximate system representations involving linear subsystems with an additional (and constant) input and only nonlinearities that take *absolute values* can be proved using Theorem 3 of Ref. 18.

The proof of Theorem 2 can easily be modified to establish a corresponding result for the discrete-time case in which Ω is replaced with $\Omega_d \triangleq \{0, 1, \cdots\}$. In fact, for that case the proof simplifies in an important conceptual way because then for any positive integer ω, $(S_1)_\omega$ is compact with respect to the usual discrete-time analog of ρ_ω in the proof of the theorem. In particular, in the discrete-time case we can set $n = p$, set $S = S_1$, and take H to be the identity map from S onto itself. This leads to the following theorem in which $\mathscr{L}_\infty(n)$ is $L_\infty(n)$ with Ω replaced with Ω_d, and $k(q)$ stands for the collection of all functions k from Ω_d^q to the $1 \times n^q$ matrices such that (11) holds for all $\tau_1, \cdots, \tau_q$ and all j where $R_j < \infty$ and the $\phi_{jir}(\tau_i)$ are real and are nonzero for at most a finite number of values of τ_i.

Theorem 3: Let $U = \{u \in \mathscr{L}_\infty(n): \|u\| \leq \beta\}$ in which β is a positive number, and let K be a map from U to the real-valued functions defined on Ω_d such that K is causal, time invariant and an element of $\mathscr{A}(U)$ in the sense of Section 2.1 with Ω replaced with Ω_d. Let K satisfy the continuity condition that given $y \in U$ and numbers $t \in \{1, 2, \cdots\}$ and $\delta_1 > 0$, there is a $\delta_2 > 0$ such that $|(Ky)(t) - (Kz)(t)| < \delta_1$ whenever $z \in U$ and $\max_{\tau \in \{0,\ldots,t\}} |y(\tau) - z(\tau)| < \delta_2$. Then, given any $\epsilon > 0$, there is a positive integer Q, and elements k_q of $k(q)$ $(1 \leq q \leq Q)$ such that

$$|(Ku)(t) - (Vu)(t)| < \epsilon, \qquad t \in \Omega_d$$

for all $u \in U$, where

$$(Vu)(t) = \sum_{q=1}^{Q} \sum_{\tau_1=0}^{t} \cdots \sum_{\tau_q=0}^{t} k_q(\tau_1, \cdots, \tau_q)\chi[u(t - \tau_1), \cdots, u(t - \tau_q)].$$

REFERENCES

1. I. W. Sandberg, "Expansions for Nonlinear Systems," B.S.T.J., *61,* No. 2 (February 1982), pp. 159–99.
2. I. W. Sandberg, "On Volterra Expansions for Time-varying Nonlinear Systems," IEEE Trans. Circuits Syst., *CAS-30* (February 1983), pp. 61–7.
3. I. W. Sandberg, "The Mathematical Foundations of Associated Expansions for Mildly Nonlinear Systems," IEEE Trans. Circuits Syst., *CAS-30* (July 1983), pp. 441–55.
4. M. Fliess, M. Lamnabhi, and F. Lamnabhi-Lagarrique, "An Algebraic Approach to Nonlinear Functional Expansions," IEEE Trans. Circuits Syst. *CAS-30* (August 1983), pp. 554–70.
5. R. J. P. de Figueiredo, "A Generalized Fock Space Framework for Nonlinear System and Signal Analysis," IEEE Trans. Circuits Syst., *CAS-30* (September 1983), pp. 637–47.
6. S. Boyd, "Volterra Series: Engineering Fundamentals," dissertation, University of California, Berkeley, March 1985.
7. I. W. Sandberg, "Criteria for the Global Existence of Functional Expansions for Input/Output Maps," AT&T Tech. J., *64,* No. 7 (September 1985), pp. 1639–58.
8. V. Volterra, *Theory of Functionals and of Integral and Integro-Differential Equations,* New York: Dover, 1959.
9. M. B. Brilliant, "Theory of the Analysis of Nonlinear Systems," MIT RLE Report, No. 345, 1958.
10. D. George, "Continuous Nonlinear Systems," MIT RLE Report, No. 355, 1959.
11. W. Porter and T. Clark, "Causality Structure and the Weierstrass Theorem," Journal of Math. Analy. Applications, *52* (November 1975), pp. 351–63.
12. W. Root, "On the Modeling of Systems for Identification, Part I: ϵ-Representations of Classes of Systems," SIAM J. Control, *13* (July 1975) pp. 927–44.
13. P. Gallman and K. S. Narendra, "Representations of Nonlinear Systems via the Stone-Weierstrass Theorem," Automatica, *12* (November 1976) pp. 619–22.
14. W. Porter, "Approximation by Bernstein Systems," Math. Syst. Theory, *11* (November 1978), pp. 259–74.
15. W. J. Rugh, *Nonlinear System Theory: The Volterra/Wiener Approach,* Baltimore: Johns Hopkins Univ. Press, 1981, pp. 34–7.
16. S. Boyd and L. O. Chua, "Fading Memory and the Problem of Approximating Nonlinear Operators with Volterra Series," Univ. of Calif. Memorandum No. UCB/ERL M84/96, November 29, 1984.
17. I. W. Sandberg, "Existence and Evaluation of Almost Periodic Steady-State Responses of Mildly Nonlinear Systems," IEEE Trans. Circuits Syst., *CAS-31* (August 1984), pp. 689–701.
18. M. H. Stone, "A Generalized Weierstrass Approximation Theorem," *Studies in Modern Analysis, Vol. 1,* R. C. Buck, ed. Englewood Cliffs: Prentice Hall, 1962.
19. N. Wiener, *The Fourier Integral and Certain of its Applications,* New York: Dover, 1933.
20. I. W. Sandberg, "Some Results on the Theory of Physical Systems Governed by Nonlinear Functional Equations," B.S.T.J., *44,* No. 5 (May–June, 1965), pp. 871–98.

APPENDIX A

Proof of the Proposition

Let any $\delta > 0$, $u \in S$, and $t \geq 0$ be given. Choose real $\delta_1 > 0$ and $\delta_2 > 0$ such that

$$2\delta_1 + 2c\delta_2 < \delta,$$

and let Δ_1 and Δ_2 be elements of Ω for which

$$|(Fy)(\tau) - Fy_{\max[0,\tau-\Delta_1]}(\tau)| < \delta_1, \qquad \tau \geq 0 \tag{16}$$

for $y \in H(S)$, and

$$|(Hv)(\tau) - Hv_{\max[0,\tau-\Delta_2]}(\tau)| < \delta_2, \qquad \tau \geq 0 \tag{17}$$

for $v \in S$. Choose $\Delta \in \Omega$ so that $\Delta > \Delta_1 + \Delta_2$, and consider

$$\phi \triangleq |(FHu)(t) - FHu_{\max[0,t-\Delta]}(t)|.$$

If $t \in [0, \Delta]$, $u = u_{\max[0,t-\Delta]}$ and $\phi = 0$. Now let $t > \Delta$. Clearly,

$$\phi = |(FHu)(t) - (FH\hat{u})(t)|,$$

where $\hat{u} = u_{(t-\Delta)}$. Using (16), we have

$$|(FHu)(t) - F(Hu)_{(t-\Delta_1)}(t)| < \delta_1,$$

and

$$|(FH\hat{u})(t) - F(H\hat{u})_{(t-\Delta_1)}(t)| < \delta_1,$$

and one finds that

$$\begin{aligned}\phi \leqslant\ & |(FHu)(t) - F(Hu)_{(t-\Delta_1)}(t) + F(Hu)_{(t-\Delta_1)}(t) \\ & - (FH\hat{u})(t) + F(H\hat{u})_{(t-\Delta_1)}(t) - F(H\hat{u})_{(t-\Delta_1)}(t)| \\ & \leqslant 2\delta_1 + c \sup_{t\in[(t-\Delta_1),t]} |(Hu)(\tau) - (H\hat{u})(\tau)|.\end{aligned}$$

By (17), $|(Hu)(\tau) - Hu_{(\tau-\Delta_2)}(\tau)| < \delta_2$ and $|(H\hat{u})(\tau) - H\hat{u}_{(\tau-\Delta_2)}(\tau)| < \delta_2$ for $\tau > \Delta_2$. Note that $\tau \geqslant (t - \Delta_1)$ and $t > \Delta \Rightarrow \tau > \Delta_2$; and that for $\tau \geqslant (t - \Delta_1)$, $Hu_{(\tau-\Delta_2)}(\tau) = H\hat{u}_{(\tau-\Delta_2)}(\tau)$. Thus,

$$\sup_{\tau\in[(t-\Delta_1),t]} |(Hu)(\tau) - (H\hat{u})(\tau)| < 2\delta_2,$$

which shows that $\phi \leqslant 2\delta_1 + 2c\delta_2$. Since this implies that $\phi < \delta$, the proposition is proved.

APPENDIX B

An Example of an Application of Theorem 2

In this Appendix we consider systems governed by the model

$$y = Nx \tag{18}$$

$$x = Av + Cy \tag{19}$$

$$w = Dv + By, \tag{20}$$

in which v is the input, w is the output, A, B, C, and D are linear operators, N is nonlinear, and x and y can be viewed as the input and output, respectively, of the nonlinear portion of the system. Models of this kind have been used in Ref. 2 and in other papers. Here we suppose that v, w, x, and y belong to $L_\infty(n)$, $L_\infty(1)$, $L_\infty(p)$, and $L_\infty(p)$, respectively, that N is memoryless and defined by $(Nx)(t) = \eta[x(t)]$ where η is a map from $\mathbb{R}^p$ to $\mathbb{R}^p$ which takes the zero element of $\mathbb{R}^p$ into itself, that η satisfies a global Lipschitz condition $|\eta(x_a) - \eta(x_b)|$

$\leqslant \gamma |x_a - x_b|$ where $|\cdot|$ is as in the definition of the norm in $L_\infty(p)$, and that A, B, C, and D, respectively, are causal time-invariant bounded linear maps from $L_\infty(n)$ to $L_\infty(p)$, $L_\infty(p)$ to $L_\infty(1)$,, $L_\infty(p)$ to $L_\infty(p)$, and $L_\infty(n)$ to $L_\infty(1)$. In particular, we assume that C has the convolution representation

$$(Cy)(t) = \int_0^t c(t - \tau)y(\tau)d\tau, \qquad t \geqslant 0$$

for $y \in L_\infty(p)$, where $c(\cdot)$ is $p \times p$ and has integrable elements [that is, has elements that are integrable on $[0, \infty)$]. The equations of a very large class of systems with a single output can be put in this form with A, B, and D convolutions whose matrix-valued kernels are either integrable, or integrable with the exception of an impulse at the origin (see Ref. 2, Appendices I and II).

B.1 Further assumptions, and approximations

Assume in the remainder of this Appendix that $(I - CN)$ is an invertible map of $L_\infty(p)$ onto $L_\infty(p)$, where I is the identity operator on $L_\infty(p)$, and that $(I - CN)^{-1}$ is causal, time invariant, and globally Lipschitz. Conditions under which these assumptions are met can be obtained from standard existence theory and results in the area of stability theory [see, for example, Ref. 20, Theorem 3 and Corollary 3(*a*)]. It follows that $w = Dv + BN(I - CN)^{-1}Av$ for all $v \in L_\infty(n)$.

With r an arbitrary positive constant, let us now restrict our inputs v to the ball $\Lambda = \{v \in L_\infty(n): \|v\| \leqslant r\}$. Let $\Lambda_1 = \{u \in L_\infty(p): \|u\| \leqslant r\|A\|\}$. In addition to the assumptions introduced above, suppose that A is a convolution with an integrable kernel, and that $(I - CN)^{-1}$, which takes Λ_1 into $L_\infty(p)$, belongs to $\mathscr{A}(\Lambda_1)$ in the sense of Section 2.1. Using the proposition in Section 2.3, and by considering one component of $(I - CN)^{-1}A$ at a time, we see that $(I - CN)^{-1}A \in \mathscr{A}(\Lambda)$, since, as can easily be verified, $A \in \mathscr{A}(\Lambda)$. Similarly, $N(I - CN)^{-1}A: \Lambda \to L_\infty(p)$ and finally $BN(I - CN)^{-1}A$ both belong to $\mathscr{A}(\Lambda)$. Thus, by Theorem 2 [with $H = A$ and $F = BN(I - CN)^{-1}$], and in the sense of Theorem 2, $BN(I - CN)^{-1}A$ can be approximated arbitrarily well on Λ by a finite Volterra series.

Before proceeding to the important matter of how one might show that $(I - CN)^{-1} \in \mathscr{A}(\Lambda_1)$ under some reasonable conditions, suppose that the assumptions described above are met, with the exception that A is *not* a convolution. Assume instead that $(Av)(t) = av(t)$, where a is a $p \times n$ matrix of constants. (This case arises naturally in the study of feedback systems.) Using the identity $(I - CN)^{-1} = (I - CN)^{-1}CN + I$, one has $w = Dv + BNAv + BN(I - CN)^{-1}CNAv$. The term $BNAv$ has a simple representation as is; and if, for example, B is a convolution

with an integrable kernel and $n = p = 1$, then, by the Weierstrass approximation theorem for real-valued continuous functions on a compact real interval, it is clear that it can be approximated arbitrarily well on Λ by a finite series having the form

$$\sum_{q=1}^{Q} \int_0^t b_q(t - \tau)v(\tau)^q d\tau, \qquad t \geqslant 0. \tag{21}$$

Consider now the more interesting term $BN(I - CN)^{-1}CNAv$. By Theorem 2 (this time with $H = C$) we see that it can be approximated arbitrarily well on Λ by a finite series of the form (11) with each $u(t - \tau_j)$ replaced with $\eta[av(t - \tau_j)]$. In particular, using the fact that the k_q in (11) satisfy

$$\int_{[0,\infty)^q} |k_{qj}(\tau_1, \cdots, \tau_q)| d(\tau_1, \cdots, \tau_q) < \infty$$

for each j, and the Weierstrass approximation theorem for real-valued continuous functions of several real variables, it follows that the term can be approximated arbitrarily well throughout Λ by a finite Volterra-like series in the sense of the sets of iterated integrals $K(m)$ in Ref. 2. [These Volterra-like series, which are frequently needed in *exact* expansion representations, can be viewed as Volterra series with symbolic kernels that include certain delta functions. A simple example of a Volterra-like series is (21).]

B.2 $(I - CN)^{-1}$ and the memory condition

The hypothesis that $(I - CN)^{-1} \in \mathscr{A}(\Lambda_1)$ plays a key role in the discussion above. We begin our comments concerning this hypothesis with the observation that with arbitrary $t \geqslant 0$, $\Delta \geqslant 0$, and $u \in \Lambda_1$, one has $[(I - CN)^{-1}u](t) - [(I - CN)^{-1}u_{\max[0,t-\Delta]}](t)$ equal to $x(t) - \tilde{x}(t)$, where x and $\tilde{x}$ are elements of $L_\infty(p)$ such that

$$x(\alpha) + \int_0^\alpha c(\alpha - \tau)\eta[x(\tau)]d\tau = u(\alpha) \tag{22}$$

$$\tilde{x}(\alpha) + \int_0^\alpha c(\alpha - \tau)\eta[\tilde{x}(\tau)]d\tau = u_{\max[0,t-\Delta]}(\alpha) \tag{23}$$

for $\alpha \geqslant 0$.

With σ any positive constant,

$$y(\alpha) + \int_0^\alpha \hat{c}(\alpha - \tau)\hat{\eta}[y(\tau), \tau]d\tau = z(\alpha) \tag{24}$$

$$\tilde{y}(\alpha) + \int_0^\alpha \hat{c}(\alpha - \tau)\hat{\eta}[\tilde{y}(\tau), \tau]d\tau = \hat{z}(\alpha) \tag{25}$$

for $\alpha \in [0, \infty)$, where $y(\alpha) = x(\alpha)e^{\sigma\alpha}$, $\tilde{y}(\alpha) = \tilde{x}(\alpha)e^{\sigma\alpha}$, $\hat{c}(\alpha) = c(\alpha)e^{\sigma\alpha}$, $\hat{\eta}[y(\tau), \tau] = e^{\sigma\tau}\eta[e^{-\sigma\tau}y(\tau)]$, $z(\alpha) = u(\alpha)e^{\sigma\alpha}$, and $\hat{z}(\alpha) = e^{\sigma\alpha}u_{\max[0,t-\Delta]}(\alpha)$. Let Σ denote the set of positive σ such that the elements of $\hat{c}$ are square integrable, and suppose that Σ is not empty. Let us now make the key assumption, which we shall refer to as A.0, that from (24) and (25) it can be concluded that for some $\sigma \in \Sigma$ there is a constant λ which depends only on c, η, and σ such that

$$\|y - \tilde{y}\|_2 \leq \lambda \|z - \tilde{z}\|_2,$$

where

$$\|y - \tilde{y}\|_2^2 = \int_0^\infty [y(t) - \tilde{y}(t)]^{Tr}[y(t) - \tilde{y}(t)]dt,$$

"Tr" denotes the transpose, and similarly for $\|z - \tilde{z}\|_2$. Much is known about conditions under which A.0 is met [see Ref. 20, Corollary 1(a) and Theorem 6], and it is known that A.0 is met in certain specific important cases.

For $t \leq \Delta$, obviously $x(t) - \tilde{x}(t) = 0$. Now let $t > \Delta$.

Notice that $\|y - \tilde{y}\|_2 \leq \xi e^{\sigma(t-\Delta)}$ where $\xi = \lambda p^{1/2} r \|A\|$. Using (24) and (25), and the Schwarz inequality,

$$|y(t) - \tilde{y}(t)| \leq \max_i \sum_{j=1}^p \left(\int_0^t |\hat{c}(\tau)_{ij}|^2 d\tau\right)^{1/2} \cdot \left(\int_0^t |\hat{\eta}_j[\tilde{y}(\tau), \tau] - \hat{\eta}_j[y(\tau), \tau]|^2 d\tau\right)^{1/2}.$$

Since

$$\int_0^t |\hat{\eta}_j[\tilde{y}(\tau), \tau] - \hat{\eta}_j[y(\tau), \tau]|^2 d\tau \leq \int_0^t |\hat{\eta}[\tilde{y}(\tau), \tau] - \hat{\eta}[y(\tau), \tau]|^2 d\tau \leq \gamma^2 \int_0^t |\tilde{y}(\tau) - y(\tau)|^2 d\tau \leq \gamma^2 \int_0^t [\tilde{y}(\tau) - y(\tau)]^{Tr}[\tilde{y}(\tau) - y(\tau)]d\tau \leq \gamma^2 \|\tilde{y} - y\|_2^2,$$

we find that

$$|y(t) - \tilde{y}(t)| \leq \xi_1 \gamma \xi e^{\sigma(t-\Delta)},$$

where

$$\xi_1 = \max_i \sum_{j=1}^{p} \left(\int_0^t |\hat{c}(\tau)_{ij}|^2 d\tau \right)^{1/2}.$$

Thus,

$$|x(t) - \tilde{x}(t)| \leqslant \xi_1 \gamma \xi e^{-\sigma\Delta}.$$

This shows that $(I - CN)^{-1} \in \mathscr{A}(\Lambda_1)$ under the conditions described, and therefore that the input-output maps of a very large class of systems have finite Volterra series approximations in the strong sense of this Appendix. [For example, using material in Ref. 20 (see the comment at the bottom of p. 875 there), it is not difficult to show that this class includes a large family of electrical networks consisting of sources, passive elements, and monotone nonlinear resistors.]

AUTHOR

Irwin W. Sandberg, B.E.E., 1955, M.E.E., 1956, and D.E.E., 1958, Polytechnic Institute of Brooklyn; AT&T Bell Laboratories, 1958—. Mr. Sandberg has been concerned with analysis of radar systems for military defense, synthesis and analysis of active and time-varying networks, with several fundamental studies of properties of nonlinear systems, and with some problems in communication theory and numerical analysis. His more recent interests have included compartmental models, the theory of digital filtering, global implicit-function theorems, and functional expansions for nonlinear systems. IEEE Centennial Medalist, Former Vice Chairman IEEE Group on Circuit Theory, and Former Guest Editor IEEE Transactions on Circuit Theory Special Issue on Active and Digital Networks. Fellow and member, IEEE; member, American Association for the Advancement of Science, Eta Kappa Nu, Sigma Xi, Tau Beta Pi, National Academy of Engineering.

PAPERS BY AT&T BELL LABORATORIES AUTHORS

COMPUTING/MATHEMATICS

Baker B. S., Coffman E. G., Willard D. E., **Algorithms for Resolving Conflicts in Dynamic Storage Allocation.** J ACM 32(2):327–343, Apr 1985.

Bentley J., **A Spelling Checker.** Comm ACM 28(5):456–462, May 1985.

Brown T. C., Erdos P., Chung F. R. K., Graham R. L., **Quantitative Forms of a Theorem of Hilbert.** J. Comb Th A 38(2):210–216, Mar 1985.

Carroll J. D., **Multidimensional Scaling—Davison, M. L. (Book Review).** Psychometri 50(1):133–140, Mar 1985.

Coffman E. G., Kadota T. T., Shepp L. A., **A Stochastic Model of Fragmentation in Dynamic Storage Allocation.** SIAM J Comp 14(2):416–425, May 1985.

Du D. Z., Hwang F. K., Yao E. Y., **The Steiner Ratio Conjecture is True for 5 Points.** J Comb Th A 38(2):230–240, Mar 1985.

Erdos P., Fowler J. C., Sos V. T., Wilson R. M., **On 2 Designs.** J Comb Th A 38(2):131–142, Mar 1985.

Foster J. C., Rosenthal C. W., Talmadge R. N., **The AT&T Technologies Unified CAD System for Electronic Design.** Comput Ind 5(4):297–309, Dec 1984.

Garey M. R., Johnson D. S., **Composing Functions to Minimize Image Size.** SIAM J Comp 14(2):500–503, May 1985.

Goldfarb R. B., Rao K. V., Chen H. S., **Differences Between Spin Glasses and Ferroglasses—Pd-Fe-Si.** Sol St Comm 54(9):799–801, Jun 1985.

Graham R. L., Winkler P. M., **On Isometric Embeddings of Graphs.** T Am Math S 288(2):527–536, Apr 1985.

Haddon R. C., Lamola A. A., **The Molecular Electronic Device and the Biochip Computer—Present Status.** P NAS US 82(7):1874–1878, Apr 1985.

Jain S. K., Agrawal V. D., **Modeling and Test-Generation Algorithms for MOS Circuits.** IEEE Comput 34(5):426–433, May 1985.

Karmarkar N., **A New Polynomial-Time Algorithm for Linear Programming.** Combinatori 4(4):373–395, 1984.

Klincewicz J. G., Luss H., **Optimal Timing Decisions for the Introduction of New Technologies.** Eur J Oper 20(2):211–220, May 1985.

Odlyzko A. M., Richmond L. B., **On the Number of Distinct Block Sizes in Partitions of a Set.** J Comb Th A 38(2):170–181, Mar 1985.

Page J. T., **Error-Rate Estimation in Discriminant Analysis.** Technomet 27(2):189–198, May 1985.

Sabnani K., Schwartz M., **Verification of a Multidestination Selective Repeat Procedure.** Comput Netw 8(5–6):463–478, Oct–Dec 1984.

ENGINEERING

Anfinsen C. B., Flory P. J., Penzias A. A., **NAS Exchange Protocol (Letter).** Science 228(4699):530, May 3 1985.

Antreasyan A., Chen C. Y., Logan R. A., **Low-Threshold InGaAsP Buried-Crescent Stop-Cleaved Lasers for Monolithic Integration.** Electr Lett 21(9):404–405, Apr 25 1985.

Baker G. L., Bates F. S., **Polyacetylene—Structure-Synthesis Relationships.** Molec Cryst 117(1–4):15–22, 1985.

Bean J. C., **Silicon MBE—From Strained-Layer Epitaxy to Device Application.** J Cryst Gr 70(1–2):444–451, Dec 1984.

Benson K. E., Lin W., **The Role of Oxygen in Silicon for VLSI.** J Cryst Gr 70(1–2):602–608, Dec 1984.

Bjorkholm J. E., Eichner L., **Monitoring the Growth of Nonuniform Gratings Written Holographically by Gaussian Laser Beams.** J Appl Phys 57(7):2402–2405, Apr 1 1985.

Bowers J. E., Koch T. L., Hemenway B. R., Wilt D. P., Bridges T. J., Burkhardt E. G., **High-Frequency Modulation of 1.52 μm Vapor-Phase-Transported InGaAsP Lasers.** Electr Lett 21(9):392–393, Apr 25 1985.

Campbell J. C., Dentai A. G., Qua G. J., Long J., Riggs V. G., **Planar InGaAs Pin Photodiode With a Semi-Insulating InP Cap Layer.** Electr Lett 21(10):447–448, May 9 1985.

Capasso F., **New Heterojunction Devices by Band-Gap Engineering.** Physica B&C 129(1–3):92–106, Mar 1985.

Chance B., Powers L., Bartunik H., Schick D., **Automatic X-Ray-Beam Position Tracking Circuit—Experimental Tests at DESY.** Rev Sci Ins 56(4):581–585, Apr 1985.

Cheung K. P., Auston D. H., **Distortion of Ultrashort Pulses on Total Internal Reflection.** Optics Lett 10(5):218–219, May 1985.

Chin A. K., **The Effect of Crystal Defects on Device Performance and Reliability.** J Cryst Gr 70(1–2):582–596, Dec 1984.

Chin A. K., Camlibel I., Caruso R., Young M. S. S., Vonneida A. R., **Effects of Thermal Annealing on Semi-Insulating Undoped GaAs Grown by the Liquid-Encapsulated Czochralski Technique.** J Appl Phys 57(6):2203–2209, Mar 15 1985.

Chiu T. H., Tsang W. T., **Reflection High-Energy Electron-Diffraction Studies on the Molecular-Beam-Epitaxial Growth of AlSb, GaSb, InAs, InAsSb, and GaLnAsSb on GaSb.** J Appl Phys 57(10):4572–4577, May 15 1985.

Chu S. N. G., Nakahara S., Strege K. E., Johnston W. D., **Surface-Layer Spinodal Decomposition in $In_{1-X}Ga_XAs_YP_{1-Y}$ and $In_{1-X}Ga_XAs$ Grown by Hydride Transport Vapor-Phase Epitaxy.** J Appl Phys 57(10):4610–4615, May 15 1985.

Doany F. E. et al., **Isomerization Intermediates in Solution Phase Photochemistry of Stilbenes.** P Soc Photo 533:25–29, 1985.

Donovan E. P., Spaepen F., Turnbull D., Poate J. M., Jacobson D. C., **Calorimetric Studies of Crystallization and Relaxation of Amorphous Si and Ge Prepared by Ion Implantation.** J Appl Phys 57(6):1795–1804, Mar 15 1985.

Dutta N. K., Cella T., Napholtz S. G., Craft D. C., **1.3 μm InGaAsP Index-Guided Multirib Wave-Guide Laser Array.** Electr Lett 21(8):326–327, Apr 11 1985.

Dutta N. K., Koszi L. A., Segner B. P., Craft D. C., Napholtz S. G., **High-Power Index-Guided Multiridge Wave-Guide Laser Array.** Appl Phys L 46(9):803–804, May 1 1985.

Dutta N. K., Napholtz S. G., Yen R., Wessel T., Shen T. M., Olsson N. A., **Long Wavelength InGaAsP (Lambda-Approximately-1.3-μm) Modified Multiquantum Well Laser.** Appl Phys L 46(11):1036–1038, Jun 1 1985.

Eisenstein G., Korotky S. K., Stulz L. W., Veselka J. J., Jopson R. M., Hall K. L., **Antireflection Coatings on Lithium-Niobate Waveguide Devices Using Electron-Beam Evaporated Yttrium Oxide.** Electr Lett 21(9):363–364, Apr 25 1985.

Fleischer P. E., Ganesan A., Laker K. R., **A Switched Capacitor Oscillator With Precision Amplitude Control and Guaranteed Start-Up.** IEEE J Soli 20(2):641–647, Apr 1985.

Forrest S. R., Kaplan M. L., Schmidt P. H., Gates J. V., **Evaluation of III-V-Semiconductor Wafers Using Nondestructive Organic-on-Inorganic Contact Barriers.** J Appl Phys 57(8):2892–2895, Apr 15 1985.

Fratello V. J., Pierce R. D., Brandle C. D., **Variation of the Temperature-Coefficient of Collapse Field in Bismuth-Based Bubble Garnets.** J Appl Phys 57(8):4043–4045, Apr 15 1985.

Friedman J. M., Ondrias M. R., Findsen E. W., Simon S. R., **Structure and Reactivity in Hemoglobin—Implications of Recent Picosecond Transient Raman and Absorption Studies.** P Soc Photo 533:8–14, 1985.

Graebner J. E., Lemaire P. J., Allen L. C., Haemmerle W. H., **Clustering of Molecular Hydrogen in Fused Silica.** Appl Phys L 46(9):839–841, May 1 1985.

Gurvitch M., Remeika J. P. Rowell J. M., Geerk J., Lowe W. P., **Tunneling, Resistive and Structural Study of NBN and Other Superconducting Nitrides.** IEEE Magnet 21(2):509–513, Mar 1985.

Gyorgy E. M., Walker L. R., **Effect of Au Additions on the Rotational Hysteresis of $Cu_{0.85}Mn_{0.15}$.** J Appl Phys 57(8):3395–3397, Apr 15 1985.

Hardwick N. E., Reynolds M. R., **Finite-Element Analysis of Thermal Stresses in the Single-Mode Bonded Splice.** Optics Lett 10(5):241–243, May 1985.

Hartney M. A., Novembre A. E., **Poly(Methylstyrene-Dimethylsiloxane) Block Copolymers as Bilevel Resists.** P Soc Photo 539:90–96, 1985.

Hebard A. F., Paalanen M. A., **Diverging Characteristic Lengths at Critical Disorder in Thin-Film Superconductors.** Phys Rev L 54(19):2155–2158, May 13 1985.

Heffner W. R., Berreman D. W., Butler G., Marcus M., **A Wedge Cell Characterization Method for the Bistable Cholesteric Twist Cell.** J Appl Phys 57(10):4507–4513, May 15 1985.

Hegarty J., Olsson N. A., McGlashan-Powell M., **Measurement of the Raman Crosstalk at 1.5 μm in a Wavelength-Division-Multiplexed Transmission System.** Electr Lett 21(9):395–396, Apr 25 1985.

Hoangbinh D., Encrenaz P., Linke R. A., **Observations of Radio Recombination Lines in the Millimeter-Wave Spectrum of Orion-A (Letter).** Astron Astr 146(1):L19–L21, May 1 1985.

Hong M., Bacon D. D., Vandover R. B., Gyorgy E. M., Dillon J. F., Albiston S. D., **Aging Effects on Amorphous Tb-Transition-Metal Films Prepared by Diode and Magnetron Sputtering.** J Appl Phys 57(8):3900–3902, Apr 15 1985.

Huggins H. A., Gurvitch M., **Preparation and Characteristics of Nb/Al-Oxide-Nb Tunnel Junctions.** J Appl Phys 57(6):2103–2109, Mar 15 1985.

Jin S., Vandover R. B., Sherwood R. C., Tiefel T. H., **Ferritic Fe-Ni Magnetic Sensor Wires With End-to-End Voltage-Generating Characteristics.** J Appl Phys 57(8):3800–3802, Apr 15 1985.

Johnson A. M., Simpson W. M., **Tunable Femtosecond Dye Laser Synchronously Pumped by the Compressed Second Harmonic of Nd-YAG.** J Opt Soc B 2(4):619–625, Apr 1985.

Johnson A. M., Simpson W. M., **Tunable Femtosecond Synchronously Modelocked Dye Laser Pumped by the Compressed Second Harmonic of Nd-YAG.** P Soc Photo 533:52–58, 1985.

Kogelnik H., **High-Speed Lightwave Transmission in Optical Fibers.** Science 228(4703):1043–1048, May 31 1985.

Koren U., Penna T. C., Tien P. K., **Heterojunction Phototransistors on N-Channelled Semi-Insulating InP Substrates.** Electr Lett 21(8):346–347, Apr 11 1985.

Korpiun P., Buchner B., Tam A. C., Wong Y. H., **Effect of Coupling Gas Viscosity on the Photoacoustic Signal.** Appl Phys L 46(11):1039–1041, Jun 1 1985.

Kunt M. et al., **Second-Generation Image-Coding Techniques.** P IEEE 73(4):549–574, Apr 1985.

Lawrence J. M., Thompson J. D., Fisk Z., Batlogg B., **Low-Temperature Resistivity of Ce-La-Th Under Pressure.** J Appl Phys 57(8):3131–3133, Apr 15 1985.

Lax M., Agrawal G. P., Belic M., Coffey B. J., Louisell W. H., **Electromagnetic-Field Distribution in Loaded Unstable Resonators.** J Opt Soc A 2(5):731–742, May 1985.

Lu C. Y., Lu N. C. C., Shih C. C., **Resistance Switching Characteristics in Polycrystalline Silicon Film Resistors.** J Elchem So 132(5):1193–1196, May 1985.

MacQueen D. B., Sannella D. T., **Completeness of Proof Systems for Equational Specifications.** IEEE Soft E 11(5):454–461, May 1985.

Magnea N., Petroff P. M., Capasso F., Logan R. A., Alavi K., Cho A. Y., **Electric-Field Dependent Cathodoluminescence of III-V-Compound Heterostructures—A New Interface Characterization Technique.** Appl Phys L 46(11):1074–1076, Jun 1 1985.

Marcatili E. A., **Dielectric Tapers With Curved Axes and No Loss.** IEEE J Q El 21(4):307–314, Apr 1985.

Miles R. O. et al., **Low-Frequency Characterization of 1.3-μ-C3 Lasers.** P Soc Photo 514:337–345, 1984.

Mollenauer L. F., Stolen R. H., Islam M. N., **Experimental Demonstration of Soliton Propagation in Long Fibers—Loss Compensated by Raman Gain.** Optics Lett 10(5):229–231, May 1985.

Ninke W. H., **Design Considerations of NAPLPS, the Data Syntax for Videotex and Teletext in North America.** P IEEE 73(4):740–753, Apr 1985.

Ogielski A. T., Morgenstein I., **Critical Behavior of Three-Dimensional Ising Model of Spin Glass.** J Appl Phys 57(8):3382–3385, Apr 15 1985.

Olgac N. M., Cooper C. A., Longman R. W., **The Impact of Costly Observations and Observation Delay in Stochastic Optimal-Control Problems.** Int J Contr 41(3):769–785, Mar 1985.

Ong E., Tai K. L., Vadimsky R. G., Kemmerer C. T., Bridenbaugh, P. M., **Structure and Composition Dependence of the Anisotropy of the Wet Chemical Etching of Germanium-Selenium Films.** P Soc Photo 539:52–55, 1985.

Ota Y., Clapper R. A., Schimmel D. G., **Thermal Behavior of Thick Vacuum Evaporated Polycrystalline Silicon Layer.** J Appl Phys 57(10):4599–4609, May 15 1985.

Pearce C. W. et al., **Oxygen Content of Heavily Doped Silicon.** Appl Phys L 46(9):887–889, May 1 1985.

Platzman P. M., Tzoar N., **Inelastic Magnetic-X-Ray Scattering.** J Appl Phys 57(8):3623–3625, Apr 15 1985.

Rapp O. et al., **Superconducting Nb-Ni Metal Glasses.** Sol St Comm 54(10):899–901, Jun 1985.

Rust R. D., **Citations Wanted (Letter).** Res Dev 27(5):183–184, May 1985.

Schmidt R. L., Haskell B. G., Eng K. Y., Oriordan S. M., **An Experimental Time-Compression System for Satellite Television Transmission.** P IEEE 73(4):789–794, Apr 1985.

Silfvast W. T., **He-Cd Laser Development (Letter).** Laser Foc E 21(5):18, May 1985.

Sooryakumar R., Chemla D. S., Pinczuk A., Gossard A. C., Wiegmann W., Sham L. J., **Valence Band Mixing in GaAs-(AlGa)As Heterostructures.** Sol St Comm 54(10):859–862, Jun 1985.

Stassis C., Axe J. D., Majkrzak C. F., Batlogg B., Remeika J., **Measurement of the Induced Moment Magnetic Form Factor of a Heavy Fermion Superconductor.** J Appl Phys 57(8):3087–3089, Apr 15 1985.

Temkin H., Panish M. B., Logan R. A., Abeles J. H., **Hybrid Growth of InGaAsP Double-Channel Planar Buried Heterostructure Lasers.** Appl Phys L 46(9):811–813, May 1 1985.

Tsang W. T., **Elimination of Oval Defects in Epilayers by Using Chemical Beam Epitaxy.** Appl Phys L 46(11):1086–1088, Jun 1 1985.

Tung R. T., Nakahara S., Boone T., **Growth of Single-Crystal $NiSi_2$ Layers on Si (110).** Appl Phys L 46(9):895–897, May 1 1985.

Vandover R. B., Hong M., Gyorgy E. M., Dillon J. F., Albiston S. D., **Intrinsic Anisotropy of Tb-Fe Films Prepared by Magnetron Co Sputtering.** J Appl Phys 57(8):3897–3899, Apr 15 1985.

Vandover R. B., Jin S., **A Simple Inductive Contactless Switch.** J Appl Phys 57(8):3798–3799, Apr 15 1985.

Varma C. M., **Heavy Fermion Superconductors and Some Associated Problems.** J Appl Phys 57(8):3064–3066, Apr 15 1985.

Wolfe R., Hegarty J., Dillon J. F., Luther L. C., Celler G. K., Trimble L. E., Dorsey C. S., **Thin-Film Wave-Guide Magneto-Optic Isolator.** Appl Phys L 46(9):817–819, May 1 1985.

MANAGEMENT/ECONOMICS

Gordon R. H., **Taxation of Corporate Capital Income—Tax Revenues Versus Tax Distortions.** Q J Econ 100(1):1–27, Feb 1985.

Ross I. M., **The Global Contest in Industrial Competitiveness Has Just Begun.** Res Manag 28(3):10–14, May–Jun 1985.

PHYSICAL SCIENCES

Appelbaum A., Knoell R. V., Murarka S. P., **Study of Cobalt-Disilicide Formation From Cobalt Monosilicide.** J Appl Phys 57(6):1880–1886, Mar 15 1985.

Barns R. L., Laudise R. A., **Size and Perfection of Crystals in Lake Ice.** J Cryst Gr 71(1):104–110, Jan–Feb 1985.

Bedeaux D. et al., **Structure of the Liquid Vapor Interface Using a Gaussian Column Model With a Variable Interaction Range.** Physica A 130(1-2):88-122, Mar 1985.

Boeshaar P. C., Tyson J. A., **New Limits on the Surface Density of M-Dwarfs. 1. Photographic Survey and Preliminary CCD Data.** Astronom J 90(5):817+, May 1985.

Brand H. R., Cladis P. E., **Physical Properties of the First Truly Ferroelectric Liquid-Crystal Phase and a Proposed Antiferroelectric Liquid-Crystal Phase.** Molec Cryst 114(1-3):207-235, 1984.

Chu S. N. G., Stevie F. A., Macrander A. T., Karlicek R. F., Chang C. C., Jodlauk C. M., Strege K. E., Mitcham D. L., Johnston W. D., **Gallium Contamination of InP Epitaxial Layers in InP/InGaAsP Multilayer Structures Grown by Hydride Transport Vapor-Phase Epitaxy.** J Elchem So 132(5):1187-1193, May 1985.

Clark W. G. et al., **Measurements of Soliton Trapping and Motion in Trans-Polyacetylene Using Dynamic Nuclear Polarization.** Molec Cryst 117(1-4):447-454, 1985.

Dewolff P. M. et al., **Nomenclature for Crystal Families, Bravais-Lattice Types and Arithmetic Classes Report of the International Union of Crystallography Ad-Hoc Committee on the Nomenclature of Symmetry.** ACT Cryst A 41(May):278-280, May 1 1985.

Dillon J. F., Rupp L. W., Batlogg B., **Spin Resonance in $Eu_xSr_{1-x}S$ With $X = 0.4$, 0.5, and 0.54.** J Appl Phys 57(8):3488-3490, Apr 15 1985.

Duncan T. M., Winslow P., Bell A. T., **The Characterization of Carbonaceous Species on Ruthenium Catalysts With C-13 Nuclear Magnetic-Resonance Spectroscopy.** J Catalysis 93(1):1-22, May 1985.

Etemad S. et al., **Band Edge and Neutral Soliton Absorption in Polyacetylene—The Role of Coulomb Correlation.** Molec Cryst 117(1-4):275-282, 1985.

Etemad S., Baker G. L., Orenstein J., Lee K. M., **Photoexcitations of a Polydiacetylene.** Molec Cryst 118(1-4):389-393, 1985.

Feldman R. D., Austin R. F., **Liquid-Phase Epitaxial Growth of InGaAsP and InP on Mesa-Patterned InP Substrates.** J Cryst Gr 71(1):1-8, Jan-Feb 1985.

Frankenthal R. P., Milner P. C., Siconolfi D. J., **Long-Term Atmospheric Oxidation of High-Purity Iron.** J Elchem So 132(5):1019-1021, May 1985.

Freund R. S. et al., **The Electronic Spectrum and Energy Levels of the Deuterium Molecule.** J Phys Ch R 14(1):235-383, 1985.

Frieze W. E., Gidley D. W., Lynn K. G., **Positron-Beam-Brightness Enhancement—Low-Energy Positron Diffraction and Other Applications.** Phys Rev B 31(9):5628-5633, May 1 1985.

Green M. L., Levy R. A., **Structure of Selective Low-Pressure Chemically Vapor-Deposited Films of Tungsten.** J Elchem So 132(5):1243-1250, May 1985.

Greenblatt M., Wang E., Eckert H., Kimura N., Herber R. H., Waszczak J. V., **Lithium Insertion Compounds of the High-Temperature and Low-Temperature Polymorphs of $LiFeSnO_4$.** Inorg Chem 24(11):1661-1665, May 22 1985.

Hamann D. R., Mattheiss L. F., **Energetics of Silicide Interface Formation.** Phys Rev L 54(23):2517-2520, Jun 10 1985.

Henley C. L., **Critical Ising Spin Dynamics on Percolation Clusters.** Phys Rev L 54(18):2030-2033, May 6 1985.

Hensel J. C., Tung R. T., Poate J. M., Unterwald F. C., **Effects of Ion Bombardment on Transport Properties of Thin Films of $CoSi_2$ and $NiSi_2$.** Nucl Inst B 7-8(Mar):409-412, Mar 1985.

Hohenberg P. C., Kramer L., Riecke H., **Effects of Boundaries on One-Dimensional Reaction Diffusion Equations Near Threshold.** Physica D 15(3):402-420, Apr 1985.

Ikedasaito M., Argade P. V., Rousseau D. L., **Resonance Raman Evidence of Chloride Binding to the Heme Iron in Myeloperoxidase.** FEBS Letter 184(1):52-55, May 6 1985.

Jayaraman A., Batlogg B., Van Uitert L. G., **Effect of High Pressure on the Raman and Electronic Absorption Spectra of $PbMoO_4$ and $PbWO_4$.** Phys Rev B 31(8):5423-5427, Apr 15 1985.

Jordan A. S., Vonneida A. R., Caruso R., **The Theory and Practice of Dislocation Reduction in GaAs and InP.** J Cryst Gr 70(1–2):555–573, Dec 1984.

Kaplan M. L., Lovinger A. J., Reents W. D., Schmidt P. H., **The Preparation, Spectral Properties, and X-Ray Structural Features of 2,3-Naphthalocyanines.** Molec Cryst 112(3–4):345–358, 1984.

Laviolette R. A., Stillinger F. H., **Consequences of the Balance Between the Repulsive and Attractive Forces in Dense, Nonassociated Liquids.** J Chem Phys 82(7):3335–3343, Apr 1 1985.

Lourenco J. A., **On the Dissolution of InGaAsP (Lambda Greater Than 1.3 μm) and InGaAs Layers by In-P Melts.** J Cryst Gr 70(1–2):155–161, Dec 1984.

MacCallum C. J., Huters A. F., Stang P. D., Leventhal M., **Search for Gamma-Ray Line Emission from SS-433.** Astrophys J 291(2):486–491, Apr 15 1985.

Martinez O. E., Fork R. L., Gordon J. P., **Theory of Passively Mode-Locked Lasers for the Case of a Nonlinear Complex-Propagation Coefficient.** J Opt Soc B 2(5):753–760, May 1985.

McWhan D. B., Aeppli G., Remeika J. P., Nelson S., **Time-Resolved X-Ray-Scattering Study of $BaTiO_3$ (Letter).** J Phys C 18(12):L307–L312, Apr 30 1985.

Meloni A., Medford L. V., Lanzerotti L. J., **Geomagnetic Anomaly Detected at Hydromagnetic Wave Frequencies.** J Geo R-S E 90(NB5):3569–3574, Apr 10 1985.

Powers L., Chance B., **Multiple Structures and Functions of Cytochrome Oxidase.** J Inorg Bio 23(3–4):207–217, Mar–Apr 1985.

Raghavachari K., **An Augmented Coupled Cluster Method and Its Application to the First-Row Homonuclear Diatomics.** J Chem Phys 82(10):4607–4610, May 15 1985.

Raghavachari K., **Basis Set and Electron Correlation Effects on the Electron Affinities of First Row Atoms.** J Chem Phys 82(9):4142–4146, May 1 1985.

Rice C. E., Jackel J. L., Brown W. L., **Measurement of the Deuterium Concentration Profile in a Deuterium-Exchanged $LiNbO_3$ Crystal.** J Appl Phys 57(9):4437–4440, May 1 1985.

Rietman E. A., **Conduction Properties of Silver and Copper Pseudohalides.** J Mat Sci L 4(5):542–544, May 1985.

Ryoji M., Yamane T., Gordon M., Kaji A., **Two Modes of Amber Codon Read-Through In Vitro.** Arch Bioch 238(2):636–641, May 1 1985.

Salamancariba L. et al., **Imaging of Defects and Recrystallization Studies in Ion-Implanted Graphite.** Nucl Inst B 7–8(Mar):487–492, Mar 1985.

Shah J., Pinczuk A., Gossard A. C., Wiegmann W., **Energy-Loss Rates for Hot Electrons and Holes in GaAs Quantum Wells.** Phys Rev L 54(18):2045–2048, May 6 1985.

Stillinger F. H., Weber T. A., **Computer Simulation of Local Order in Condensed Phases of Silicon.** Phys Rev B 31(8):5262–5271, Apr 15 1985.

Teo B. K., **Molecular-Orbital Justification of Topological Electron-Counting Theory.** Inorg Chem 24(11):1627–1638, May 22 1985.

Treacy M. M. J., Gibson J. M., Howie A., **On Elastic Relaxation and Long Wavelength Microstructures in Spinodally Decomposed $In_XGa_{1-X}As_YP_{1-Y}$ Epitaxial Layers.** Phil Mag A 51(3):389–417, Mar 1985.

Vaidya S., Retajczyk T. F., Knoell R. V., **Effect of Dopant Implantation on the Properties of $TaSi_2$ Poly-Si Composites.** J Vac Sci B 3(3):846–852, May–Jun 1985.

Vianden R., Winand P. M. J., Kaufmann E. N., MacDonald J. R., Jackmann T. E., **Precise Lattice Site Determination for Th and U Implanted Into Be Single Crystals.** Nucl Inst B 7–8(Mar):109–112, Mar 1985.

Wayda A. L., Dye J. L., **A Versatile System for Vacuum-Like Manipulations.** J Chem Educ 62(4):356–359, Apr 1985.

Widom A. et al., **Vacuum Spin Waves Through the Chiral Anomaly.** Phys Lett A 108(8):377–378, Apr 22 1985.

Winters H. F., Chang R. P. H., Mogab C. J., Evans J., Thornton J. A., Yasuda H., **Coatings and Surface Modification Using Low-Pressure Non-Equilibrium Plasmas.** Mater Sci E 70(1–2):53–77, Apr 1985.

Wolfe A. et al., **Dependence of Hydromagnetic Energy-Spectra Near $L = 2$ and $L = 3$ on Upstream Solar-Wind Parameters.** J Geo R-S P 90(NA6):5117+, Jun 1 1985.

Woolery G. L., Walters M. A., Suslick K. S., Powers L. S., Spiro T. G., **Alternative Fe-O-2 Bond Lengths in O-2 Adducts of Iron Porphyrins—Implications for Hemoglobin Cooperativity.** J Am Chem S 107(8):2370–2373, Apr 17 1985.

Yurke B., **Optical Back-Action-Evading Amplifiers.** J Opt Soc B 2(5):732–738, May 1985.

SOCIAL AND LIFE SCIENCES

Carroll J. D., Desarbo W. S., **Two-Way Spatial Methods for Modeling Individual Differences in Preference.** Adv Consum 12:571–576, 1985.

Frase L. T., **Flexibility in Writing Style—A New Discourse-Level Cloze Test.** Writ Comm 2(2):107–127, Apr 1985.

Friedman J. M., **Structure, Dynamics, and Reactivity in Hemoglobin.** Science 228(4705):1273–1280, Jun 14 1985.

Sagi D., Julesz B., **Detection Versus Discrimination of Visual Orientation.** Perception 13(5):619–628, 1984.

SPEECH/ACOUSTICS

Adkins J. M., Sorkin R. D., **Effect of Channel Separation on Earphone-Presented Tones, Noise, and Stereophonic Material.** J Aud Eng S 33(4):234–239, Apr 1985.

Flanagan J. L., Kubli R. A., **Conference Microphone With Adjustable Directivity.** J Acoust So 77(5):1946–1949, May 1985.

Perkell J. S., Nelson W. L., **Variability in Production of the Vowels (I) and (A).** J Acoust So 77(5):1889–1895, May 1985.

Sessler G. M., West J. E., **A Simple Second-Order Toroid Microphone.** Acustica 57(4–5):193–199, Apr–May 1985.

CONTENTS, NOVEMBER 1985

ANALYTICOL

AT&T TECHNICAL JOURNAL is abstracted or indexed by *Abstract Journal in Earthquake Engineering, Applied Mechanics Review, Applied Science & Technology Index, Chemical Abstracts, Computer Abstracts, Current Contents/Engineering, Technology & Applied Sciences, Current Index to Statistics, Current Papers in Electrical & Electronic Engineering, Current Papers on Computers & Control, Electronics & Communications Abstracts Journal, The Engineering Index, International Aerospace Abstracts, Journal of Current Laser Abstracts, Language and Language Behavior Abstracts, Mathematical Reviews, Science Abstracts (Series A, Physics Abstracts; Series B, Electrical and Electronic Abstracts; and Series C, Computer & Control Abstracts), Science Citation Index, Sociological Abstracts, Social Welfare, Social Planning and Social Development,* and *Solid State Abstracts Journal.* Reproductions of the Journal by years are available in microform from University Microfilms, 300 N. Zeeb Road, Ann Arbor, Michigan 48106.


AT&T