Partial-Response Signaling

PETER KABAL AND SUBBARAYAN PASUPATHY, MEMBER, IEEE

Abstract—This paper presents a unified study of partial-response signaling (PRS) systems and extends previous work on the comparison of PRS schemes. A PRS system model is introduced which enables the investigation of PRS schemes from the viewpoint of spectral properties such as bandwidth, nulls, and continuity of derivatives. Several desirable properties of PRS systems and their relation to system functions are indicated and a number of useful schemes, some of them not previously analyzed, are presented. These systems are then compared using as figures of merit speed tolerance, minimum eye width, and signal-to-noise ratio (SNR) degradation over ideal binary transmission. A new definition of speed tolerance, which takes into account multilevel outputs and the effect of sampling time, is introduced and used in the calculation of speed-tolerance figures. It is shown that eye width, a performance measure that has not been used previously in comparing PRS systems, can be calculated analytically in many cases. Exact values as well as bounds on the SNR degradation for the systems under consideration are presented. The effect of precoding on system performance is also analyzed.

I. INTRODUCTION

PULSE-AMPLITUDE modulation (PAM) is often used to convey digital information. The usual constraint on permissible PAM signal waveforms is that they should not cause intersymbol interference. Signal design based on this criterion can sometimes lead to a complete intolerance of timing errors or to incompatibilities with some channel characteristics. Some of these disadvantages can be removed with partial-response signaling (PRS) (also known as correlative level coding) wherein the constraint on waveforms is relaxed so as to allow a controlled amount of intersymbol interference. PRS designs are based on the premise that since the intersymbol interference is known, its effect can be removed.

One of the merits of PRS is that the controlled intersymbol interference can be used to shape the system spectrum, for instance to place nulls in the frequency response. Also, this spectral shaping can make the system less sensitive to timing errors. This allows practical PRS systems to transmit at the Nyquist rate, a feat not possible with ordinary PAM. In addition, a PRS spectrum might be chosen to complement a nonideal characteristic in order to reduce the residual undesired intersymbol interference. The partial-response coding format has the further advantage that violations in the code can be used

The authors are with the Department of Electrical Engineering, University of Toronto, Toronto, Ont., Canada.

to monitor error performance or even to correct errors. On the negative side, PRS systems using symbol by symbol detection possess reduced noise margins due to the fact that the superposition of signal waveforms causes the number of output levels to be larger than the number of input levels.

Lender [1] first introduced duobinary PRS as a data-transmission method. Kretzmer [2] categorized the characteristics of several PRS schemes and compared them on the basis of speed tolerance and signal-to-noise ratio (SNR) degradation. PRS techniques have been applied to various modulation schemes such as FM [3] and SSB [4] as well as to baseband systems. Several authors [5]–[7] have investigated error detection and control for PRS. Lender introduced precoding for duobinary [1]. Gerrish and Howson [8] discussed precoding for multilevel inputs. Tomlinson [9] and Harashima and Miyakawa [10] have introduced a generalized form of PRS precoding. Kobayashi [11], in addition to comparing PRS with other coding techniques, has provided a good bibliography. Maximum-likelihood decoding for PRS has also been an active area of research [12], [13]. The special problems associated with adaptive equalization for PRS have also been examined [14], [15].

This paper, partly tutorial in nature, presents a unified study of PRS and extends previous work on the comparison of PRS schemes. First a cohesive framework for studying PRS systems is introduced. To this end, a general scheme which separates PRS waveform generation into two parts is described in Section II. This allows the investigation of PRS spectra from the viewpoint of useful properties such as nulls and continuity of derivatives.

In Section III a number of candidate PRS systems are selected by choosing ones with desirable spectral properties and a small number of output levels. The PRS schemes under consideration include several whose performances have not been assessed previously. The process of decoding and precoding and the effect of error propagation are then discussed in Section IV. The rest of the paper is devoted to a comparison of PRS systems using three criteria. The sensitivity to timing errors is measured by two parameters, speed tolerance and the minimum eye width, while the effect of an increase in the number of output levels is measured by the SNR degradation over ideal binary transmission. Speed-tolerance figures are calculated using a new definition of speed tolerance which takes into account multilevel systems and sampling phase. The eye-width criterion has not been used previously for comparing PRS systems. A procedure is given here which

greatly simplifies the calculation of eye widths. Comparative SNR degradation figures are also presented. These include exact values for systems without precoding, figures not previously reported.

II. A GENERALIZED PRS SYSTEM

The baseband model of a synchronous data communications system is shown in Fig. 1. The model is applicable to various modulation schemes. The overall transfer function $H(\omega)$ encompasses the transmitter filter (or signal generator characteristic), the equivalent baseband channel, and the receiving filter (which may include an equalizer).

First the discussion will concentrate on an ideal, noiseless system (ideal in that there is no distortion due to channel imperfections or sampler offsets). Such a system can be characterized by the samples of the desired impulse response $h(t)$. Let N be the smallest number of contiguous samples that span all the nonzero samples. Then if $\{f_n\}$, $n = 0,1,2,\cdots,N - 1$ are these N sample values, the PRS system polynomial is given by

$$F(D) = \sum_{n=0}^{N-1} f_n D^n \qquad (1)$$

where D is the delay operator. For a given input symbol sequence $\{x_n\}$ the output sequence $\{y_n\}$ is given by

$$Y(D) = X(D)F(D) \qquad (2)$$

where

$$X(D) = \sum_{n=0}^{\infty} x_n D^n \quad \text{and} \quad Y(D) = \sum_{n=0}^{\infty} y_n D^n.$$

The $\{x_n\}$ will be assumed to be independent m-ary symbols taking on the equally likely values $\{-(m-1), -(m-3), \cdots, (m-3), (m-1)\}$.

Though most of the desirable properties of a PRS system can be stated in terms of the impulse response $h(t)$, many are best illuminated by frequency domain considerations. Fig. 2 shows a method of generating the PRS system function $H(\omega)$ which gives an insight into frequency domain properties.

The system consists of a tapped delay line with coefficients $\{f_n\}$ in cascade with a filter with frequency response $G(\omega)$. The transversal filter has the periodic frequency response (period $2\pi/T$)

$$\mathfrak{F}(\omega) = F(D)|_{D=\exp(-j\omega T)}$$

$$= \sum_{n=0}^{N-1} f_n \exp\ (-j\omega n T) \qquad (3)$$

where T is the symbol spacing. It can be easily shown that $h(t)$ has the sample values $\{f_n\}$ if and only if $G(\omega)$ satisfies Nyquist's first criterion,[1] that is

---

[1] Nyquist's first criterion requires that all but one of the samples of the impulse response be zero, i.e., $g(kT) = 0$, $k \neq 0$, $g(0) = 1$ [16].
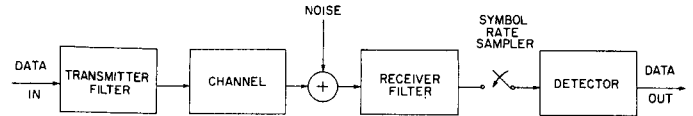

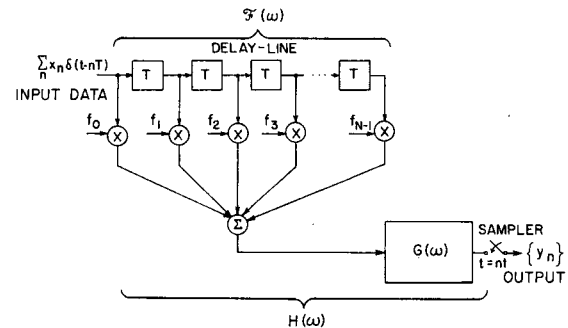Fig. 1.   Model for a synchronous data communications system.

Fig. 2.   General partial-response system model.

$$\sum_{k=-\infty}^{\infty} G\left(\omega - \frac{2\pi k}{T}\right) = T. \qquad (4)$$

The PRS system has been separated into two parts: $\mathfrak{F}(\omega)$ forces the desired sample values but is periodic, while $G(\omega)$ preserves the sample values but may be used to bandlimit the resulting system function. For a given system polynomial $F(D)$, different choices of $G(\omega)$ satisfying Nyquist's first criterion result in different system functions $H(\omega)$, all of which have identical sampled responses. The advantage of this separation will be more apparent when we study the various desirable properties of $H(\omega)$ and see how some of these properties (such as spectral nulls) are influenced by the choice of the sample values $\{f_n\}$ and by the choice of $G(\omega)$. The separation is of course an artifice—the actual implementation may be organized quite differently. While Fig. 2 naturally suggests a digital partial-response encoder (i.e., $\mathfrak{F}(\omega)$), it may be advantageous in some cases to implement the total system function $H(\omega)$ in an analog fashion.

III. CHOOSING THE PRS SYSTEM POLYNOMIAL

A. System Bandwidth

In order to maximize the data rate in the available bandwidth, many PRS systems are designed to occupy the minimum bandwidth which supports transmission without undesired intersymbol interference, i.e., $H(\omega) = 0$ for $|\omega| > \pi/T$. Such systems also avoid the aliasing which occurs when the output of a nonminimum bandwidth system is sampled at the symbol rate. This aliasing can in extreme cases cause dips in the middle of the Nyquist equivalent channel which are difficult to equalize.

For minimum bandwidth systems (see (4)),

$$G(\omega) = \begin{cases} T, & |\omega| \leq \pi/T \\ 0, & \text{elsewhere.} \end{cases} \qquad (5)$$

The corresponding system impulse response is given by

$$h(t) = \sum_{n=0}^{N-1} f_n \frac{\sin \frac{\pi}{T}(t - nT)}{\frac{\pi}{T}(t - nT)}. \qquad (6)$$

Other choices for $G(\omega)$ (occupying a larger bandwidth) are possible and, as we shall see, allow the use of system polynomials (such as $1 - D$) which are unsuitable for minimum bandwidth systems.

B. Spectral Null at $\omega = \pi/T$

It is well known that if $H(\omega)$ and its first $K - 1$ derivatives are continuous and the Kth derivative is discontinuous, $|h(t)|$ decays asymptotically as $1/|t|^{K+1}$ [17]. Continuity of the function and its derivatives helps to reduce the portion of the total signal energy in the tails of $h(t)$ and hence the undesirable intersymbol interference in nonideal systems. For minimum bandwidth systems, $\mathfrak{F}(\omega)$ must have a zero at π/T (where $G(\omega)$ has a discontinuity) for $H(\omega)$ to be a continuous function. In general the conditions for the continuity of $H(\omega)$ and its derivatives are summed up as follows.

Proposition 1: The first $K - 1$ derivatives of a minimum bandwidth $H(\omega)$ are continuous if and only if $F(D)$ has $(1 + D)^K$ as a factor.

If $F(D)$ has more than one zero at $D = -1$, the roll-off near $\omega = \pi/T$ becomes less sharp and thus the design of practical filters for the system becomes easier. On the other hand, if $F(D)$ has a large multiplicity of $1 + D$ factors, the error performance tends to be degraded due to the increase in the number of output levels. The sensitivity to timing offsets also suffers because more controlled intersymbol interference terms are introduced. These affect the timing sensitivity more than the distant tails which decay rapidly as more factors $1 + D$ are introduced.

For systems that are not restricted to the minimum bandwidth, a null at π/T is still useful—a pilot tone inserted at this point can be used for clock recovery [18].

C. Spectral Null at $\omega = 0$

Reduced low-frequency components in the spectrum are desirable in systems such as transformer coupled circuits, dc powered cables, SSB modems, and carrier systems with carrier pilot tones. For a null at $\omega = 0$, it can be seen from (3) that $1 - D$ must be a factor of $F(D)$. Multiple zeros at $D = 1$ cause a more gradual roll-off of the frequency components just above dc which may also be desirable.

With combinations of just the two factors $1 + D$ and $1 - D$, most of the common partial-response systems can be developed. Table I shows a number of PRS system polynomials and the corresponding $|H(\omega)|$ and $h(t)$.

TABLE I
PARTIAL-RESPONSE SYSTEMS FORMED FROM $1 + D$ AND $1 - D$

F (D)	|H(ω)|	h(t)
$1+D$ (duobinary-class 1)		
$1-D$ (dicode)		
$(1+D)(1-D)$ $= 1-D^2$ (modified duobinary-class 4)		
$(1+D)^2$ $= 1+2D+D^2$ (class 2)		
$(1+D)^2(1-D)$ $= 1+D-D^2-D^3$		
$(1+D)(1-D)^2$ $= 1-D-D^2+D^3$		
$(1+D)^2(1-D)^2$ $= 1-2D^2+D^4$ (class 5)		
$(1+D)(2-D)$ $= 2+D-D^2$ (class 3)		
$(1+D)(1-D)(2+D^2)$ $= 2-D^2-D^4$		

The class designations in the first column of Table 1 are due to Kretzmer [2]. Expressions for $H(\omega)$ and $h(t)$ are given in Table II. In the general formula for $h(t)$, (6), the sample f_0 occurs at time zero; however a change of time origin often simplifies the resulting expression. In Table II, the center of the nonzero samples has been chosen as the time origin, that is in (6), $t = (N - 1)T/2$ is the new time origin. The expressions for the frequency response in Table II also assume this time origin.

The first entry in these tables, duobinary, also satisfies Nyquist's second criterion, i.e., that the pulsewidth should be undistorted. The second system $1 - D$ is not practical in the minimum bandwidth because of the resulting discontinuity in the system function $H(\omega)$. Modified duobinary, the next entry, has both a dc null and a null at $\omega = \pi/T$. The fourth entry, $1 + 2D + D^2$, has the same response as a full raised cosine characteristic but sampled at twice the usual rate. Also included in the tables are two pulse responses which have not been analyzed before in the literature, namely $1 + D - D^2 - D^3$ and $1 - D - D^2 + D^3$. These two systems have both dc nulls and nulls at $\omega = \pi/T$. The last two entries in Table I suggest that

TABLE II
CHARACTERISTICS OF MINIMUM BANDWIDTH PARTIAL-RESPONSE SYSTEMS

SYSTEM POLYNOMIAL F(D)	FREQUENCY RESPONSE $H(\omega)$ for $|\omega| \leq \pi/T$	IMPULSE RESPONSE $h(t)$	NO. OF OUTPUT LEVELS L
$1 + D$	$2T \cos \frac{\omega}{2} T$	$\frac{4T^2}{\pi} \frac{\cos(\pi t/T)}{T^2 - 4t^2}$	$2m-1$
$1 - D$	$j2T \sin \frac{\omega}{2} T$	$\frac{8T}{\pi} \frac{t \cos(\pi t/T)}{4t^2 - T^2}$	$2m-1$
$1 - D^2$	$j2T \sin \omega T$	$\frac{2T^2}{\pi} \frac{\sin(\pi t/T)}{t^2 - T^2}$	$2m-1$
$1 + 2D + D^2$	$4T \cos^2 \frac{\omega}{2} T$	$\frac{2T^3}{\pi t} \frac{\sin(\pi t/T)}{T^2 - t^2}$	$4m-3$
$1 + D - D^2 - D^3$	$j4T \cos \frac{\omega T}{2} \sin \omega T$	$-\frac{64T^3 t}{\pi} \frac{\cos(\pi t/T)}{(4t^2-9T^2)(4t^2-T^2)}$	$4m-3$
$1 - D - D^2 + D^3$	$-4T \sin \frac{\omega T}{2} \sin \omega T$	$\frac{16T^2}{\pi} \frac{\cos(\pi t/T)(4t^2-3T^2)}{(4t^2-9T^2)(4t^2-T^2)}$	$4m-3$
$1 - 2D^2 + D^4$	$-4T \sin^2 \omega T$	$\frac{8T^3}{\pi t} \frac{\sin(\pi t/T)}{t^2 - 4T^2}$	$4m-3$
$2 + D - D^2$	$T + T \cos \omega T + j3T \sin \omega T$	$\frac{T^2}{\pi t} \sin(\pi t/T)(\frac{3t-T}{t^2-T^2})$	$4m-3$
$2 - D^2 - D^4$	$-T + T \cos 2\omega T + j3T \sin 2\omega T$	$\frac{2T^2}{\pi t} \sin(\pi t/T)(\frac{2T-3t}{t^2-4T^2})$	$4m-3$

any appropriate polynomial in D may be used to modify the basic polynomials $1 \pm D$.

If the system bandwidth is allowed to increase beyond π/T, the Nyquist filter $G(\omega)$ need not have discontinuities. In this case, it is permissible for $\mathfrak{F}(\omega)$ to be nonzero at $\omega = \pi/T$. Fig. 3 shows an example of a nonminimum bandwidth PRS spectrum for a pulse response $1 - D$ and a raised cosine Nyquist filter. For this example it is the Nyquist filter $G(\omega)$ which controls the spectral properties at the high end, while the system polynomial $F(D)$ controls the low-end response.

D. The Number of Output Levels

A PRS system with M nonzero pulse samples will have m^M output levels for m-ary input unless there are special relationships between the sample values. The number of output levels L lies in the range

$$M(m-1) + 1 \leq L \leq m^M \qquad (7)$$

with the minimum value being obtained when the pulse samples have the same magnitude. The number of output levels for a practical PRS system is limited both by the complexity of implementation and the inevitable distortions present in real systems. In addition there is a tendency for the error performance at a given SNR to degrade with a large number of output levels. It can easily be shown that if the system polynomial has $1 \pm D$ as a factor, some of the output levels coalesce. Thus these factors are also desirable from the viewpoint of reducing the number of output levels.

If we restrict the number of output levels to be less than 5 for binary inputs (multilevel inputs will result in considerably more output levels), the task of searching for candidate systems approaches manageability. If in addition we require that $1 + D$ and/or $1 - D$ be a factor of $F(D)$ and that there be no nulls or severe ripples in the
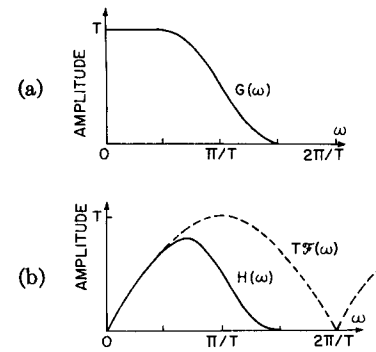

Fig. 3. Partial-response system using a raised cosine filter. (a) Raised cosine filter (50 percent excess bandwidth). (b) Partial-response spectrum for $1 - D$ pulse response.

middle of the passband, an investigation of the possible systems shows that all the suitable ones have already been listed in Tables I and II. For binary input these systems have either 3 or 5 output levels.

IV. DECODING AND PROBABILITY OF ERROR

A. Decoding

The output of an ideal noiseless system is given by

$$y_n = f_0 x_n + \sum_{i=1}^{N-1} f_i x_{n-i}. \qquad (8)$$

The receiver can recover the data x_n by subtracting out the effect of previous input symbols. In practice the receiver makes estimates of the data and then uses these to cancel the tails of the pulse response. The detector structure (Fig. 4) is the same as that of a decision feedback equalizer [19]; indeed the tasks of decoding and equalization for deviations from the desired pulse response could be combined.

When additive noise is present, the input to the slicer (Fig. 4) is given by

$$r_n = \frac{1}{f_0} \left(y_n + \zeta_n - \sum_{i=1}^{N-1} f_i \hat{x}_{n-i} \right) \qquad (9)$$

where ζ_n is a noise sample. r_n is then quantized to the nearest allowable data level to give the next data estimate $\hat{x}_n$. If we define a decision error as $e_n = x_n - \hat{x}_n$, then from (8) and (9),

$$r_n = x_n + \frac{1}{f_0} \left(\zeta_n + \sum_{i=1}^{N-1} f_i e_{n-i} \right). \qquad (10)$$

We see from (10) that past decision errors can adversely affect subsequent decisions. A technique for preventing this error propagation phenomenon, namely precoding, will be discussed later.

B. Probability of Error (No Precoding)

The probability of (symbol) error for the system shown in Fig. 4 is given by[2]

[2] For Gray coded input, the symbol error rate is approximately $\log_2 m$ times the bit error rate.
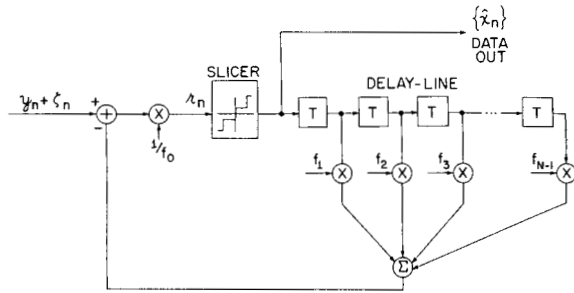
Fig. 4. Partial-response decoder (no precoding).

$$P_e = \left(1 - \frac{1}{m}\right) P\left[\left|\zeta_n + \sum_{i=1}^{N-1} f_i e_{n-i}\right| > f_0\right] \quad (11)$$

where for simplicity f_0 is assumed to be positive. This expression is difficult to evaluate in its full generality. However, with suitable assumptions, some results can be obtained.

A lower bound on P_e can be found by ignoring error propagation, i.e., setting $e_n = 0$. Then if the noise is Gaussian, the lower bound is

$$P_{eL} = 2\left(1 - \frac{1}{m}\right) Q(f_0/\sigma) \quad (12)$$

where σ^2 is the variance of ζ_n and

$$Q(x) \triangleq \frac{1}{(2\pi)^{1/2}} \int_x^\infty \exp(-u^2/2) \, du. \quad (13)$$

For white Gaussian noise Duttweiler, Mazo, and Messerschmitt [20] have given a simple upper bound on P_e for binary inputs. This bound, extended to consider m-ary input symbols, is

$$P_{eU} = \frac{m^{N-1} P_{eL}}{\dfrac{m}{m-1} P_{eL}(m^{N-1} - 1) + 1}. \quad (14)$$

Equation (14) shows that error propagation increases the error probability by at most a factor m^{N-1}. For small m and N this effect is modest.

If the noise is white, the exact probability of error can be obtained for small m and N by modeling the system as a Markov chain [21] with l^{N-1} states where l is the number of distinct values the error e_n can assume ($l = 2m - 1$). Analytically for $m = 2$ and $N = 3$,

and

$$A_2 = A_1 + \tfrac{1}{2}Q\left(\frac{f_0 + 2f_1}{\sigma}\right)\left[1 + \tfrac{1}{2}Q\left(\frac{f_0 - 2f_1 - 2f_2}{\sigma}\right)\right.$$
$$\left. - \tfrac{1}{2}Q\left(\frac{f_0 - 2f_1 + 2f_2}{\sigma}\right)\right] + \tfrac{1}{2}Q\left(\frac{f_0 - 2f_1}{\sigma}\right)$$
$$\cdot \left[1 + \tfrac{1}{2}Q\left(\frac{f_0 + 2f_1 - 2f_2}{\sigma}\right) - \tfrac{1}{2}Q\left(\frac{f_0 + 2f_1 + 2f_2}{\sigma}\right)\right].$$
$$(15)$$

For a larger number of states, an analytic solution becomes unwieldy and a numerical approach becomes more appealing.

Increases in the probability of error due to the propagation of errors (i.e., P_e/P_{eL}) for the various PRS systems were calculated and appear in Table III. The values have essentially reached their asymptotic values at $P_{eL} = 10^{-5}$. It can be seen that the effect of error propagation can increase drastically with the number of input levels in some cases.

C. Equivalent Systems

The probability of error remains unchanged for certain transformations of system polynomials (assuming uncorrelated noise samples). Kobayashi's proof of the duality between $1 + D$ and $1 - D$ [12] can be extended to show that the systems $F(D)$ and $F(-D)$ have identical error rates.

The systems $F(D)$ and $F(D^k)$ also have equivalent performances since the latter is a k-fold interleaved version of the former. Then the system polynomial may for the purposes of error analysis be reduced in order, if the greatest common divisor of the exponents of D is greater than unity. This order reduction tightens the upper bound (14) and greatly reduces the number of states in the Markov chain model.

The PRS systems under consideration fall into the following four groups, with all responses within a group having the same error performance. (These equivalences do not carry over to timing sensitivity.)

1) $1 + D, 1 - D, 1 - D^2$
2) $1 + 2D + D^2, 1 - 2D^2 + D^4$
3) $2 + D - D^2, 2 - D^2 - D^4$

$$P_e = \frac{A_2 Q(f_0/\sigma)}{A_2 Q\left(\dfrac{f_0}{\sigma}\right) + A_1\left[1 + Q\left(\dfrac{f_0}{\sigma}\right) - \tfrac{1}{2}Q\left(\dfrac{f_0 + 2f_2}{\sigma}\right) - \tfrac{1}{2}Q\left(\dfrac{f_0 - 2f_2}{\sigma}\right)\right]}$$

where

$$A_1 = \left[1 - \tfrac{1}{2}Q\left(\frac{f_0 + 2f_1 + 2f_2}{\sigma}\right)\right]\left[1 - \tfrac{1}{2}Q\left(\frac{f_0 - 2f_1 + 2f_2}{\sigma}\right)\right]$$
$$- \tfrac{1}{4}Q\left(\frac{f_0 - 2f_1 - 2f_2}{\sigma}\right) Q\left(\frac{f_0 + 2f_1 - 2f_2}{\sigma}\right)$$

4) $1 + D - D^2 - D^3, 1 - D - D^2 + D^3. \quad (16)$

D. Precoding

Precoding is used to alleviate the error-propagation problems of the previous decoder. The precoder eliminates

TABLE III
INCREASE IN PROBABILITY OF ERROR DUE TO ERROR PROPAGATION

SYSTEM	$P_{eL} = 10^{-2}$			$P_{eL} = 10^{-5}$		
	m=2	4	8	m=2	4	8
$1 + D$, $1 - D$, $1 - D^2$	1.9	3.8	7.1	2.0	4.0	8.0
$1 + 2D + D^2$, $1 - 2D^2 + D^4$	3.7	11	28	4.0	13	43
$2 + D - D^2$, $2 - D^2 - D^4$	1.9	3.7	7.1	2.0	4.0	8.0
$1 + D - D^2 - D^3$, $1 - D - D^2 + D^3$	4.5	16	41	5.0	21	96

the effect of previous symbols at the source where they are known precisely. Consider a new data sequence $\{w_n\}$ defined by[3,4]

$$w_n = \frac{1}{f_0}\Big[x_n - \sum_{i=1}^{N-1} f_i w_{n-i}\Big]. \qquad (17)$$

However w_n can take on more than m values. By interpreting w_n modulo m, the redundancy in the values of w_n is eliminated. The coefficients $\{f_n\}$ must then be integer valued. For convenience let them be normalized so that they have a greatest common divisor of unity. The precoded symbols $\{w_n\}$ are the m-ary values which satisfy

$$w_n f_0 = x_n - \sum_{i=1}^{N-1} f_i w_{n-i} \,(\text{mod } m). \qquad (18)$$

A solution exists and is unique if and only if f_0 and m are relatively prime [22]. Equation (18) can be modified slightly if $f_0, \cdots, f_{l-1}$ are zero modulo m and f_l is relatively prime to m,

$$w_{n-l} f_l = x_n - \sum_{i=l+1}^{N-1} f_i w_{n-i} \quad (\text{mod } m). \qquad (19)$$

This form is useful for PRS systems for which (18) is not applicable, e.g., $2 + D - D^2$ with binary input.

Equations (18) or (19) can be rearranged to become

$$x_n = \sum_{i=0}^{N-1} f_i w_{n-i} \quad (\text{mod } m). \qquad (20)$$

This indicates that in the absence of noise the original data can be recovered by interpreting the received signal modulo m, a memoryless operation. A real (noisy) system would in addition employ a slicer to resolve the channel output to the nearest allowable level.

A precoder also acts to some extent as a scrambler. The source data may contain long runs or have some periodic subsequences. The precoder tends to break these up, especially in systems with the more complicated system polynomials.

The statistics of the precoded sequence $\{w_n\}$ are needed to evaluate system performance. It can be shown that if the input data $\{x_n\}$ are equally likely and statistically independent, the precoded sequence will also have equiprobable independent symbols [23]. In other words, precoding replaces one data sequence by another with the same statistics. The equally likely nature of the precoded data will be used in the calculation of the probability of error; their independence will be needed for computing the SNR degradation.

The restrictions imposed on (19) mean that not all PRS systems can be precoded (though precoding is possible in most circumstances for the systems under consideration here). A generalized precoding scheme [9], [10] can be formulated for any PRS system, specifically the coefficients $\{f_n\}$ need not be integers. The precoder has a form similar to (18) but with a modulo operation defined on real numbers. The precoded digits are not in general integers and span a range 0 to m (cf., 0 to $m-1$ for ordinary precoding). The ordinary precoder is implemented easily with digital logic, while the generalized form needs much more complicated circuitry. Any references to precoding in the sequel will mean ordinary precoding although some of the results carry over to generalized precoding. A major obstacle to the analysis of systems using generalized precoding is that the statistics of the precoded sequence are difficult to obtain.

E. Probability of Error with Precoding

In order to keep the error rate expressions simple, it will be assumed that the output levels are evenly spaced. The output levels of a PRS system are not equiprobable if any of the levels coalesce.[5] However, the outer levels always have probabilities of $1/m^M$. The probability of error of a precoded system is then approximately

$$P_e \lessapprox 2(1 - 1/m^M)Q(d/\sigma), \qquad (21)$$

where σ^2 is the noise variance at the decoder and d is the decision distance (half the level separation, equal to unity for all of the specific examples considered here). Since the output data are taken modulo m from the slicer output, errors which carry the slicer output m levels away may still be correct modulo m. The probability of error (21) is an approximation in that this effect is ignored.

F. Error Detection and Optimal Decoding

The output of a PRS system has a built-in redundancy since the number of output levels is larger than the

---

[3] The precoder formulation follows Gerrish and Howson [8].
[4] The alphabet $\{0,1,\cdots,m-1\}$ will be used in this subsection to simplify the subsequent equations which involve modulo arithmetic.

[5] For $M = 2$, $L = 3$, the outputs have probabilities $1/4$, $1/2$, and $1/4$; for $M = 3$, $L = 5$, the probabilities are $1/8$, $1/4$, $1/4$, $1/4$, and $1/8$; and for $M = 4$, $L = 5$, the probabilities are $1/16$, $1/4$, $3/8$, $1/4$, and $1/16$.

number of input levels. This redundancy can be used to detect errors and hence monitor performance [7].

Maximum-likelihood sequence estimation is a decoding technique which makes full use of the correlations between signal levels at different sampling times [12], [13]. Much of the penalty imposed by symbol-by-symbol decoding of PRS is thus recovered. This improved performance is, however, at the expense of a substantially increased complexity and a decoding delay.

V. SPEED TOLERANCE

One of the reasons for using PRS is that it can allow signaling at the Nyquist rate. Ordinary PAM signaling is not practical at this rate since the resulting system is intolerant of timing perturbations. Some measures are then needed to gauge the timing sensitivity of a particular system or to compare one system with another. Speed tolerance and minimum eye width are two such figures of merit. While speed tolerance measures the sensitivity of a system to changes in the signaling rate, eye width is a measure of sensitivity to changes in the sampler phase. (It is assumed that when the rate changes, the synchronous nature of the system allows the receiver to sample the output at the changed rate.)

A. Speed Tolerance and Peak Eye Closure

Speed-tolerance figures for some PRS systems appear in the literature [2]. They seem to have been based on a definition of speed tolerance as the increase in transmission rate at which the peak eye closure just becomes unity [24, p. 90].

Lucky, Salz, and Weldon [24] offer two definitions for three-level systems, the zero level, and outer level peak eye closures. These definitions depend on specific choices for the decision thresholds. Corresponding definitions for more than three output levels have not been given and it is not clear from the literature what criterion has been used in such cases to determine the speed-tolerance figures.

The basic problem with the above definitions is that a unity value for the peak eye closure merely indicates that the distortion has exceeded an arbitrary threshold and not whether the eye pattern is open or closed. Also, when the transmission rate is no longer the nominal value, the problem of finding the "best" sampling phase must also be considered in calculating speed tolerance.

B. Speed Tolerance and Zero Eye Opening

In order to avoid the ambiguities associated with the previous definitions of speed tolerance, we will consider the conditions under which a noiseless system "fails." The corresponding rate will then define the speed tolerance.

Most PRS systems operate such that some of the m^M output levels coalesce. When the signaling rate is larger than the design value (with the transmission characteristic being kept fixed), the output levels generally split into m^M levels again. In addition, intersymbol interference from samples that are nominally zero (at the nominal transmission rate and nominal sampling phase) will cause the
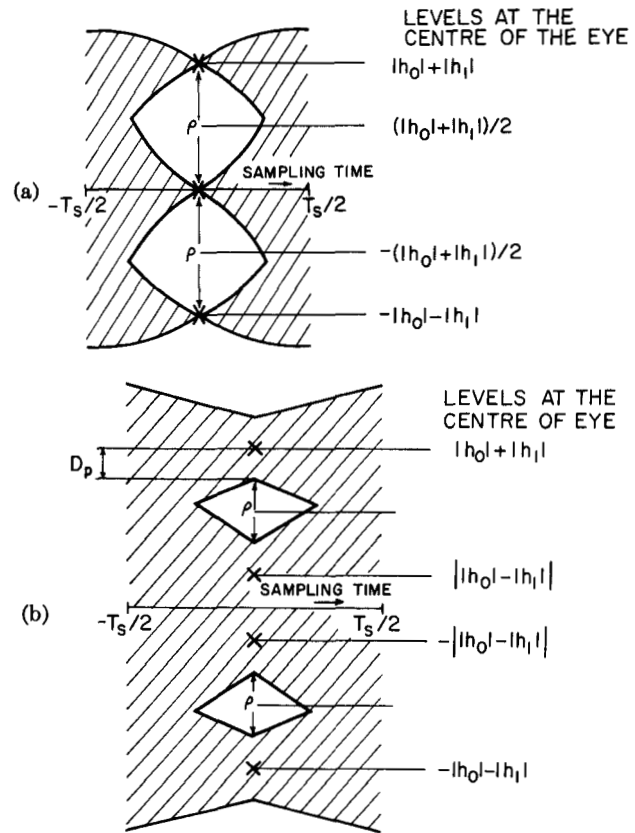

Fig. 5. Received levels for a three-level partial-response system. (a) Undistorted reception. (b) Distorted reception.

levels to spread out around these m^M levels[6] (Fig. 5).

If the decision thresholds are placed midway in the gaps between the levels[7] and the sampler phase is kept at the point where the eye opening is the largest, a noiseless system will not make errors until the levels overlap in one of the eyes (i.e., the eye closes). In general, the threshold positions as well as the sampler phase will have to be constantly readjusted as the signaling rate is varied, in order to keep the thresholds in the middle of the gaps and the sampling instant at the point of maximum eye height. Above the rate at which one of the eyes closes, some data sequences will cause errors even in the absence of noise. When this happens, we will deem the system to have failed. For other threshold locations, the system will fail at even lower transmission rates in the absence of noise. Thus it is clear that this rate is an upper bound on permissible transmission rate, the actual thresholds will dictate how close to this maximum rate the speed may approach.

We shall define speed tolerance as the increase in transmission rate at which the smallest eye opening is zero (the eye opening being the maximum height of an eye). The

---

[6] For equiprobable bipolar data, the probability of the output is symmetrical about zero. In addition, the probability distribution of the unwanted intersymbol interference is symmetric about each of the m^M levels individually.

[7] It should be noted that this placement of the thresholds does not minimize the average probability of error in the presence of Gaussian noise.

distortion from the nominally zero pulse samples at time $t_0 + nT_s$ is given by

$$\sum_{\substack{i=-\infty \\ i \notin \{I_{NZ}\}}}^{\infty} x_{n-i} h(t_0 + iT_s) \leq (m-1) \sum_{\substack{i=-\infty \\ i \notin \{I_{NZ}\}}}^{\infty} | h(t_0 + iT_s) |$$

$$\triangleq D_p, \qquad (22)$$

where T_s is the sampling interval (not necessarily equal to the nominal symbol spacing T), t_0 represents the sampler phase ($| t_0 | \leq T_s/2$), $\{I_{NZ}\}$ is the set of subscripts i for which $h(iT) \neq 0$, and D_p is the peak (or worst case) distortion from the nominally zero pulse samples.

For a three-level system (only possible with binary inputs), the eye opening is given by (Fig. 5)

$$\rho = | h_0 | + | h_1 | - || h_0 | - | h_1 || - 2D_p, \qquad (23)$$

where h_0 and h_1 are samples of the impulse response corresponding to the two nonzero pulse coefficients. The condition $\rho = 0$ can be rewritten as

$$\frac{D_p}{(| h_0 | + | h_1 | - || h_0 | - | h_1 ||)/2} = 1. \qquad (24)$$

This is analogous to the peak eye closure criteria [24].

C. Inner and Outer Eye Openings

For systems with more than three output levels, speed tolerance is derived from the signaling rate which causes at least one of the eyes to be closed.

Consider a five-level PRS system with three nonzero pulse coefficients. Let the corresponding impulse response samples be h_0, h_1, and h_2 ordered such that $| h_0 | \geq | h_1 | \geq | h_2 |$. The positive received levels due to just these samples, for binary ± 1 input, are shown in Fig. 6.

The outer eye opening is

$$\rho_O = (| h_0 | + | h_1 | + | h_2 | - D_p)$$
$$\quad - (| h_0 | + | h_1 | - | h_2 | + D_p)$$
$$= 2 | h_2 | - 2D_p, \qquad (25)$$

while the inner eye opening is

$$\rho_I = (| h_0 | - | h_1 | + | h_2 | - D_p)$$
$$\quad - (|| h_0 | - | h_1 | - | h_2 || + D_p)$$
$$= \begin{cases} 2 | h_2 | - 2D_p, & \text{for } | h_0 | \geq | h_1 | + | h_2 | \\ 2 | h_0 | - 2 | h_1 | - 2D_p, & \text{for } | h_0 | \leq | h_1 | + | h_2 |. \end{cases}$$
$$(26)$$

From (25) and (26) we see that

$$\rho_I \leq 2 | h_2 | - 2D_p = \rho_O. \qquad (27)$$

Similarly it can be shown that for all five-level systems, the inner eye opening is always less than or equal to the outer opening. Then for binary input, the speed tolerance of the systems under consideration is the increase in transmission rate which causes the central eye to close.
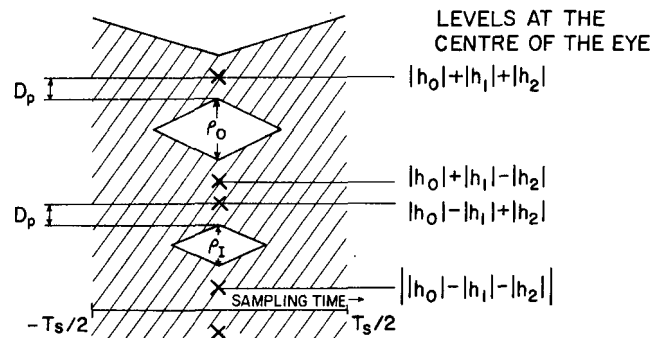

Fig. 6. Received levels for a five-level partial-response system.

D. Conditions for Nonzero Speed Tolerance

Before calculating the speed tolerance for various PRS systems on the basis of the zero eye opening criterion, the conditions under which the speed tolerance is nonzero will be determined. From the definitions of eye openings, it is clear that speed tolerance is nonzero only if D_p is bounded.

For a minimum bandwidth PRS system the expression for $h(t)$, (6), can be substituted into (22) to give

$$D_p = \frac{T(m-1)}{\pi} \sum_{\substack{i=-\infty \\ i \notin \{I_{NZ}\}}}^{\infty} \left| \sin \frac{\pi}{T}(t_0 + iT_s) \sum_{n=0}^{N-1} \frac{(-1)^n f_n}{t_0 + iT_s - nT} \right|.$$

$$(28)$$

It can be shown that the series (28) defining the peak distortion D_p converges if and only if [23]

$$1) \quad \sum_{n=0}^{N-1} (-1)^n f_n = 0 \qquad (29)$$

and/or

$$2) \quad t_0 = kT \quad \text{and} \quad T_s = lT, \quad \text{for } k, l \text{ integers.} \qquad (30)$$

The condition (29) is equivalent to requiring that $1 + D$ be a factor of the system polynomial $F(D)$. This requirement is not surprising in light of the discussion in Section III on how $1 + D$ affects the asymptotic decay of the impulse response. The second condition (30) is that the sampling instants should occur at the nominal points. These conditions lead to the conclusion that for minimum bandwidth systems, if $1 + D$ is not a factor of the system polynomial, the speed tolerance is zero since any change in speed $(1/T_s)$ will cause the peak distortion to be unbounded.

E. Comparison of PRS Systems on the Basis of Speed Tolerance

Speed tolerance figures for PRS systems of interest were calculated on a digital computer. The problem was programmed as a two-variable optimization; the speed was increased to close the eye while the sampler phase was adjusted to maximize the eye opening. This procedure was repeated until an equilibrium was reached.

Table IV shows the speed-tolerance figures for the various systems under consideration for binary inputs.

TABLE IV
SPEED TOLERANCE AND MINIMUM EYE WIDTH FOR MINIMUM BANDWIDTH PRS WITH BINARY INPUTS

SYSTEM	SPEED TOLERANCE %	MINIMUM EYE WIDTH
$1 + D$	42.5	0.667T
$1 + 2D + D^2$	26.6	0.689T
$2 + D - D^2$	7.43	0.243T
$1 - D^2$	15.5	0.357T
$1 - 2D^2 + D^4$	7.42	0.249T
$1 + D - D^2 - D^3$	14.1	0.363T
$1 - D - D^2 + D^3$	4.81	0.200T
$2 - D^2 - D^4$	2.96	0.164T

Comparison with figures previously published by Kretzmer [2][8] shows significant differences in two cases, namely $1 + 2D + D^2$ and $2 + D - D^2$.

One of the systems that has not been previously analyzed, $1 + D - D^2 - D^3$, has nearly the same speed tolerance as the well known $1 - D^2$ system. The two systems are closely related, but $1 + D - D^2 - D^3$ has an additional null at the Nyquist frequency which makes the high-end roll-off more gradual.

F. Effect of Sampler Phase

The optimum sampler phase remained unchanged with respect to the center of the impulse response for all the systems with symmetry (or antisymmetry) in their impulse responses. This is not unexpected since the eye patterns for these systems are symmetrical about the center of the impulse response. Fig. 7 shows how the zero eye opening criterion relates to the sampler phase and excess speed for the two systems with unsymmetrical responses. The optimum times were offset by $-2.7 \times 10^{-2}T$ and $-2.9 \times 10^{-2}T$ from their nominal points for the systems $2 + D - D^2$ and $2 - D^2 - D^4$, respectively. The speed tolerances for these systems are 7.4 and 3 percent, respectively. For $2 + D - D^2$, the often-quoted speed tolerance is 38 percent [2]. Fig. 7(a) shows clearly that both pairs of eyes are completely closed at transmission rates more than 8 percent over the nominal rate.

G. Nonminimum Bandwidth Systems

As noted in Section III-C, PRS is possible with systems not having $1 + D$ as a factor of the system polynomial if the system bandwidth is allowed to increase beyond the Nyquist bandwidth. As an example, we will consider using the $1 - D$ response in conjunction with a Nyquist filter having a gradual cutoff. The particular filter chosen belongs to the class of raised cosine filters [24]

[8] Kretzmer's speed-tolerance figures for $1 + D$, $1 + 2D + D^2$, $2 + D - D^2$, $1 - D^2$, and $1 - 2D^2 + D^4$ are 43, 40, 38, 15, and 8 percent, respectively.
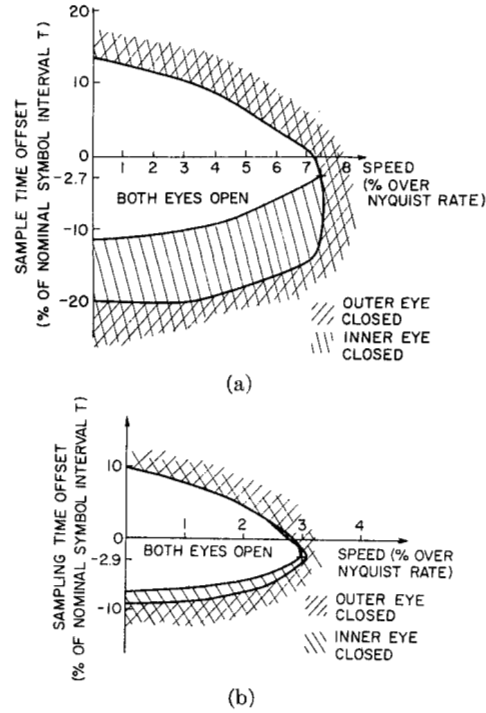

Fig. 7. Eye closure as a function of sampling time and transmission rate. (a) $2 + D - D^2$. (b) $2 - D^2 - D^4$.

$$G(\omega) = \begin{cases} T, & 0 \leq |\omega| \leq \pi(1 - \alpha)/T \\ \dfrac{T}{2}\left\{1 - \sin\left(\dfrac{T}{2\alpha}\left[|\omega| - \pi/T\right]\right)\right\}, & \\ & \pi(1 - \alpha)/T \leq |\omega| \leq \pi(1 + \alpha)/T \\ 0, & \pi(1 + \alpha)/T \leq |\omega| \end{cases}$$

(31)

where α is the roll-off parameter, $0 \leq \alpha \leq 1$. Fig. 3 illustrates the spectrum of the $1 - D$ response used with a raised cosine filter ($\alpha = 1/2$). Fig. 8 shows how the speed tolerance of this system varies with α. For comparison, this graph also shows the curves for the binary $1 + D$ and $1 - D^2$ responses. At $\alpha = 1/2$ the $1 - D$ system has a speed tolerance of 20 percent. This is better than any of the minimum bandwidth systems considered earlier which also have dc nulls. This improved speed tolerance is of course at the expense of more bandwidth; however, the tradeoff may be useful. This example serves to illustrate that Nyquist filters other than the minimum bandwidth one can be useful for generating PRS spectra.

H. Speed Tolerance for a General PRS System

The calculation of the eye opening which is used to arrive at the speed-tolerance figure can be generalized for PRS systems other than those considered in detail here. For a general PRS system, the output due just to the nominally nonzero pulse samples is given by (cf. (22))

$$u(t_0 + nT_s) = \sum_{i \in \{I_{NZ}\}} x_{n-i}h(t_0 + iT_s). \quad (32)$$

For a fixed value of $t_0 + nT_s$, only M input digits contribute to the sum. The worst-case excursions of each of
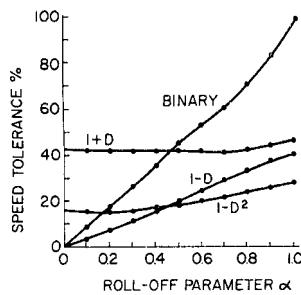
Fig. 8. Speed tolerance as a function of the roll-off parameter of a raised cosine Nyquist filter.

Fig. 9. Duobinary eye pattern.

the total of m^M possible values can be calculated by adding $\pm D_p$ to each of the m^M values. (Note that D_p varies with t_0, see (22).) As t_0 is varied and the M x_n's which affect (32) take on all combinations of values, the worst case excursions trace out the outline of the eye pattern.

The nominal output levels result when $t_0 = 0$ and $T_s = T$. For each of these nominal levels there will be corresponding M-tuples of input symbols $\{x_{n-i}\}$, $i \in \{I_{nz}\}$. The gap corresponding to a particular eye can be found by taking the difference between the minimum of the values of $u(t_0 + nT_s) \pm D_p$ caused by M-tuples which correspond to the nominal level above the eye and the maximum of the values caused by M-tuples which correspond to the nominal level below the eye. For the general case, several· eye openings must be examined, since the inner eye is not necessarily the smallest.

To help clarify the procedure, consider, for example, the duobinary eye pattern for binary input symbols (Fig. 9). The four 2-tuples $\{1,1\}$, $\{1,-1\}$, $\{-1,1\}$, and $\{-1,-1\}$ correspond to the. nominal output levels 2, 0, 0, and -2, respectively. Consider the upper eye; the lower eye boundary, due to the $\{1,-1\}$ or $\{-1,1\}$ input pair, is given by

$$\max \left[h(t_0) - h(t_0 + T_s), -h(t_0) + h(t_0 + T_s) \right] + D_p$$
(33)

where $h(t)$ is the duobinary response given in Table II. The upper boundary, in this case due only to the $\{1,1\}$ input symbol pair, is given by

$$h(t_0) + h(t_0 + T_s) - D_p.$$
(34)

The eye opening may then be calculated as the difference between (34) and (33).

VI. EYE WIDTH

To simplify the calculation of the eye width, a new formula for peak distortion is first developed.

A. Peak Distortion

To measure eye width, only the sampler phase t_0 is varied; the symbol spacing T_s assumes the nominal value T. Thus, the peak distortion due to the nominally zero pulse samples is given by (28) with $T_s = T$. Minimum bandwidth systems have bounded D_p and hence nonzero eye widths only if $1 + D$ is a factor of the system poly-
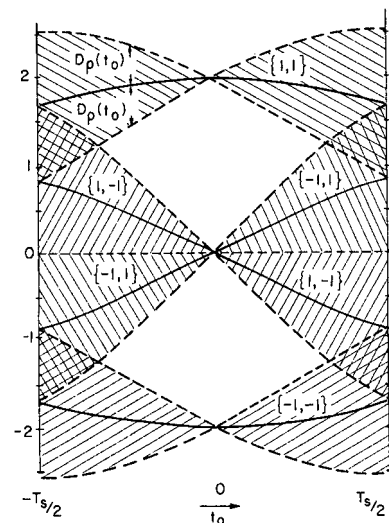
nomial (see (29) and (30)). For the rest of the section, we shall assume that such is the case.

The impulse response of a general PRS system (6) can be written as

$$h(t) = \frac{T}{\pi} \sin (\pi t/T) \sum_{n=0}^{N-1} \frac{(-1)^n f_n}{t - nT}.$$
(35)

The sine factor inserts the regularly spaced zero crossings. The second factor (i.e., the sum) has at most $N - 1$ roots and hence at most $N - 1$ sign changes. Let T_L and T_U be the smallest and largest real roots, respectively.[9] The peak distortion can be split into three terms with the real roots of the second factor of (35) contained within the range of the middle term,

$$D_p = (m - 1)\{ \sum_{i=-\infty}^{I_L} |h(t_0 + iT)| + \sum_{\substack{i=I_L+1 \\ i \notin \{I_{NZ}\}}}^{I_U-1} |h(t_0 + iT)|$$

$$+ \sum_{i=I_U}^{\infty} |h(t_0 + iT)|\}$$
(36)

where I_L is the greatest integer i satisfying $t_0 + iT \leq T_L$ and $i \leq -1$ and I_U is the smallest integer k satisfying $t_0 + kT \leq T_U$ and $k \geq N$.

Expanding the first term of (36) using (35) and then interchanging the summation and magnitude operations,

$$\sum_{i=-\infty}^{I_L} |h(t_0 + iT)|$$

$$= \frac{T}{\pi} \sin (\pi |t_0|/T) \sum_{i=-\infty}^{I_L} \left| \sum_{n=0}^{N-1} \frac{(-1)^n f_n}{t_0 + (i-n)T} \right|$$

$$= \frac{T}{\pi} \sin (\pi |t_0|/T) \left| \sum_{i=-\infty}^{I_L} \sum_{n=0}^{N-1} \frac{(-1)^n f_n}{t_0 + iT} \right.$$

$$\left. - \sum_{n=1}^{N-1} \sum_{i=I_L-n+1}^{I_L} \frac{(-1)^n f_n}{t_0 + iT} \right|.$$
(37)

[9] If there are no real roots, T_L and T_U may be set to zero to be consistent with the sequel.

The first part of the result (37) is zero since $1 + D$ is a factor of the system polynomial. Similar manipulation of the last term of (36) is possible. Then for minimum bandwidth systems with bounded peak distortion, the peak distortion can be written as a finite series,

$$D_p = \frac{(m-1)T}{\pi} \sin (\pi \mid t_0 \mid /T) \left\{ \left| \sum_{n=1}^{N-1} \sum_{i=I_L-n+1}^{I_L} \frac{-(-1)^n f_n}{t_0 + iT} \right| \right.$$
$$+ \sum_{\substack{i=I_L+1 \\ i \notin \{I_{NZ}\}}}^{I_U-1} \left| \sum_{n=0}^{N-1} \frac{(-1)^n f_n}{t_0 + (i-n)T} \right|$$
$$+ \left. \left| \sum_{n=1}^{N-1} \sum_{i=I_U-n}^{I_U-1} \frac{(-1)^n f_n}{t_0 + iT} \right| \right\}. \tag{38}$$

For all the specific PRS systems introduced previously T_L and T_U lie in the interval $[0,(N-1)T]$ giving $I_L = -1$ and $I_U = N$ in (38).

B. Eye Width

The eye opening as a function of t_0 can be calculated by adding $\pm D_p$ to the levels generated by the M nominally nonzero pulse samples. The width of the eyes can be calculated from the values of t_0 at which the eyes just close.

As a performance measure for PRS systems, we will use the smallest of the widths of the eyes. In the simpler cases, the M-tuples of input which result in the boundaries of the smallest eye can be determined analytically.[10] The upper and lower edges of the eye can be equated to give a polynomial in t_0 from whose roots the eye width can be determined.

To illustrate the procedure, the duobinary $(1 + D)$ eye width will be determined. From (38),

$$D_p = \frac{2(m-1)}{\pi} \frac{\sin (\pi \mid t_0 \mid /T)}{1 - (t_0/T)^2}. \tag{39}$$

From (32) the output from just the two main pulses is

$$\frac{\sin (\pi t_0/T)}{(\pi t_0/T)(1 - (t_0/T)^2)} \left(\frac{t_0}{T} (x_n + x_{n-1}) + x_n - x_{n-1} \right). \tag{40}$$

Fig. 9 shows these levels as a function of t_0 for binary valued x_n's. If $t_0 = 0$, (40) takes on the value $x_n - x_{n-1}$. The term $x_n + x_{n-1}$ adds distortion when $t_0 \neq 0$. It can be verified that this distortion is worst for the inner eyes for m-ary input symbols. The upper boundary of the (upper) inner eye is generated by

$$x_n = \begin{cases} -(m-3), & t_0 \geq 0 \\ m-1, & t_0 < 0 \end{cases}$$
$$x_{n-1} = \begin{cases} -(m-1), & t_0 \geq 0 \\ m-3, & t_0 < 0. \end{cases} \tag{41}$$

[10] Smith [25] has found the eye width for modified duobinary $(1 - D^2)$ in this manner. More recently work by Craig [26] has come to our attention.

The lower outline is generated by

$$x_n = x_{n-1} = \begin{cases} (m-1), & t_0 \geq 0 \\ -(m-1), & t_0 < 0. \end{cases} \tag{42}$$

The upper and lower outlines of the smallest eye then are

$$2[(2-m)\mid t_0 \mid /T + 1] \frac{\sin (\pi \mid t_0 \mid /T)}{(\pi \mid t_0 \mid /T)(1 - (t_0/T)^2)} - D_p \tag{43}$$

and

$$\frac{2(m-1) \sin (\pi \mid t_0 \mid /T)}{\pi(1 - (t_0/T)^2)} + D_p. \tag{44}$$

Substituting for D_p from (39) and equating (43) and (44), we can solve for the two values of t_0 at which the edges of the eye meet. The eye width is then just the difference between them,

$$\text{eye width} = \frac{2T}{4m - 5}. \tag{45}$$

The results of Section V-C can be applied directly— for binary input and the systems under consideration the central eye widths are the smallest. The minimum eye widths are given in Table IV.

The speed tolerance and eye width criteria have some common ground in that they measure the sensitivities to timing perturbations, albeit not of the same type. The results show that the systems with nearly equal speed tolerances have nearly equal eye widths. The major change in ordering of the PRS systems on the basis of these two measures occurs with the interchange of $1 + D$ and $1 + 2D + D^2$, the system $1 + 2D + D^2$ has the best eye width while $1 + D$ has the best speed tolerance.

VII. SNR DEGRADATION

In this section, the SNR degradation over ideal binary transmission is used to gauge the system performance with additive white Gaussian noise present. We will again concentrate on minimum bandwidth PRS.

A. System Model

In order to calculate the SNR degradation, we must specify how the filtering is apportioned between the transmitter and the receiver. Two system models are shown in Fig. 10. The arrangement in Fig. 10(a) (which will be called model 1) optimally distributes the shaping between the transmitter and receiver for a perfect channel. The phase response of the system function may be arbitrarily distributed since it does not affect the system performance in the presence of noise. The noise at the decoder is correlated at the sample times meaning that the expressions for error probability which assume independence of noise samples at the detector cannot be applied for model 1. However the lower bound (12) still holds.

In Fig. 10(b) (model 2), the transmitter filter and the channel determine the shape of the frequency response
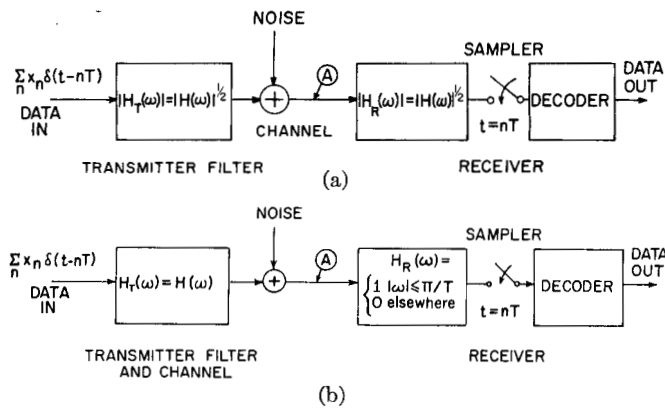
Fig. 10.   Partial-response system models. (a) Model 1. (b) Model 2.

while the receiver filter merely bandlimits the noise. The noise samples will be independent since the noise has been bandlimited to the Nyquist frequency π/T. Thus the expressions for the exact error rate as well as the bounds can be applied to model 2.

The error bounds (12) and (14) as well as the error rate with precoding (21) can be expressed by the general form

$$P_E = KQ(d/\sigma) \qquad (46)$$

where the value of K is suitably chosen, σ^2 is the noise variance at the detector, and d is the decision distance (either f_0 or half the signal level separation). The general form (46) is applicable even for exact error rates in systems without precoding provided K includes the relative increase in error rate due to error propagation which results in a given value of P_E.

The noise variance at the detector is given by

$$\sigma^2 = \frac{N_0}{2\pi} \int_{-\infty}^{\infty} |H_R(\omega)|^2 \, d\omega \qquad (47)$$

where N_0 is the noise power spectral density and $H_R(\omega)$ is the frequency response of the receiver filter (different for models 1 and 2).

B. SNR

The SNR of interest is at the receiver input (point A in Fig. 10(a) and (b)) and is given by

$$\eta = \frac{\dfrac{\sigma_x^2}{2\pi T} \displaystyle\int_{-\infty}^{\infty} |H_T(\omega)|^2 \, d\omega}{N_0/T} \qquad (48)$$

where σ_x^2 is the variance of the input symbols $[\sigma_x^2 = (m^2 - 1)/3]$, and N_0/T is the noise power in the Nyquist band $[-\pi/T, \pi/T]$.

The SNR required to achieve a given probability of error P_E can be obtained from (46)–(48),

$$\eta = \left[\frac{\sigma_x}{2\pi d} Q^{-1}(P_E/K)\right]^2 \int_{-\infty}^{\infty} |H_R(\omega)|^2 \, d\omega \int_{-\infty}^{\infty} |H_T(\omega)|^2 \, d\omega \qquad (49)$$

where $Q^{-1}(p)$ is the solution x_p of $p = Q(x_p)$.

For model 1, (49) becomes

$$\eta_1 = \left[\frac{\sigma_x}{2\pi d} Q^{-1}(P_E/K) \int_{-\infty}^{\infty} |H(\omega)| \, d\omega\right]^2. \qquad (50)$$

For model 2, using Parseval's relationship and (6), the SNR can be written as

$$\eta_2 = \left[\frac{\sigma_x}{d} Q^{-1}(P_E/K)\right]^2 \sum_{n=0}^{N-1} f_n^2. \qquad (51)$$

The PRS system will be compared with ideal binary signaling which has an SNR (see (12))

$$\eta_b = [Q^{-1}(P_E)]^2. \qquad (52)$$

The SNR degradation will then be η_1/η_b or η_2/η_b for models 1 and 2, respectively.

We shall now investigate how certain factors such as error rate, number of output levels, and precoding affect the SNR degradation.

1) *Equivalent Systems:* The transformations of the system polynomials, $F(D)$ to $F(-D)$ and $F(D)$ to $F(D^k)$, do not change the SNR degradation values for models 1 and 2. (This can be shown easily by substituting (5) and (3) into (50) or (51).) Thus the equivalences introduced in Section IV-C for error rate apply to SNR degradation also (see (16)).

2) *Precoding:* When precoding is applied to a PRS system, the decision distance d may decrease from the value it assumes with no precoding. This occurs if f_0 is not the smallest magnitude pulse coefficient (e.g., $2 + D - D^2$ and $2 - D^2 - D^4$). This decrease in decision distance means that the SNR degradation will increase when precoding is applied to such systems.

3) *Probability of Error:* We also note from (50)–(52) that the SNR degradation varies with the error rate, unless K in (46) is equal to unity. The effect of K on the SNR degradation vanishes for low error rates, i.e., $Q^{-1}(P_E/K) \to Q^{-1}(P_E)$ as $P_E \to 0$. Hence the SNR degradation becomes asymptotically independent of P_E,

$$\frac{\eta_1}{\eta_b} \geq \lim_{P_E \to 0} \frac{\eta_1}{\eta_b} = \left[\frac{\sigma_x}{2\pi d} \int_{-\infty}^{\infty} |H(\omega)| \, d\omega\right]^2 \qquad (53)$$

$$\frac{\eta_2}{\eta_b} \geq \lim_{P_E \to 0} \frac{\eta_2}{\eta_b} = \frac{\sigma_x^2}{d^2} \sum_{n=0}^{N-1} f_n^2. \qquad (54)$$

4) *Number of Output Levels:* Generally, one feels that an increase in the number of output levels should result in an increase in the SNR degradation. However (54) indicates that this need not be so and in fact indicates how to choose a PRS system with a small SNR degradation. Consider the class of PRS systems with system polynomial

$$F_k(D) = k + D - D^2 + \cdots - (-D)^k, \qquad k = 1, 2, \cdots. \qquad (55)$$

$F_k(D)$ has $1 + D$ as a factor and has $2k + 1$ output levels. With no precoding, the decision distance for this

TABLE V
SNR DEGRADATION FOR BINARY INPUTS

System	No. of Output Levels	SNR DEGRADATION (dB) at $P_E = 10^{-5}$			
		Model 1		Model 2	
		No Precoding (Lower Bound)	Precoded	No Precoding*	Precoded
1+D, 1-D, 1-D²	3	2.1	2.3	3.0 - 3.3 3.3	3.2
1 + 2D + D² 1 - 2D + D⁴	5	6.0	6.3	7.8 - 8.4 8.4	8.0
2 + D - D² 2 - D² - D⁴	5	1.2	7.5	1.8 - 2.4 2.1	8.0
1 + D - D² - D³ 1 - D - D² + D³	5	4.6	4.9	6.0 - 6.9 6.7	6.3

* The three figures in each entry represent the lower and upper bounds and the actual SNR degradation

class of systems is $d = k$. Hence the asymptotic SNR degradation from (54) is

$$\lim_{P_E \to 0} \frac{\eta_2}{\eta_b} = \sigma_x^2 \left(1 + \frac{1}{k}\right). \tag{56}$$

For large k (and hence large number of output levels) this limit approaches the lower bound on SNR degradation for m-ary input symbols. (The lower bound is achieved by ordinary PAM.) This class of systems demonstrates that SNR degradation need not necessarily increase with an increase in the number of output levels.

The effect of error propagation on the SNR degradation (as manifested in the value of K) will also tend to be small for this class as compared to other systems with the same number of pulse coefficients (for $k > 1$). This is due to the fact that the feedback coefficients are small relative to the main pulse sample on which the decision is based. On the other hand, when precoding is applied to systems in this class, the decision distance decreases to unity. The SNR degradation then increases with k,

$$\frac{\eta_2}{\eta_b} \geq \lim_{P_E \to 0} \frac{\eta_2}{\eta_b} = \sigma_x^2 (k^2 + k). \tag{57}$$

Two of the PRS systems under consideration here fall into the category specified by (55), $1 + D$ ($k = 1$) and $2 + D - D^2$ ($k = 2$).

C. Comparison on the Basis of SNR Degradation

Table V lists the SNR degradation in decibels for PRS systems with binary inputs and at the representative error rate of 10^{-5}. Kretzmer [2] lists values of SNR degradation for some of the systems considered here using model 1 and the asymptotic formula (53). Qureshi and Newhall [27] use a discrete channel model which corresponds to model 2 for SNR degradation calculations for the same systems Kretzmer considered.

Table V lists a lower bound on the SNR degradation, corresponding to the lower bound on the error rate (12), for model 1. For model 2, the SNR degradation corresponding to the bounds on the probability of error, (12) and (14), as well as the exact SNR degradation from the Markov chain model are given. For precoded systems the degradation values corresponding to the error rate of (21) are given for both system models.

Previous results have included only lower bounds on the SNR degradation for systems without precoding and thus it has not been possible to ascertain how seriously the error-propagation phenomenon affects the system performance. The results given here (for binary input) show that precoding does not decrease the SNR degradation significantly and in fact increases it for some systems. (For $2 + D - D^2$ and $2 - D^2 - D^4$ a 6-dB penalty is incurred.)

The error-propagation phenomenon is more serious for multilevel input, though the worst-case system in Table II ($m = 8$, $P_e/P_{eL} = 96$) requires only 1.6 dB more SNR without precoding than the same system precoded to maintain an error rate of 10^{-5}.

VIII. CONCLUSIONS

Five useful PRS systems have been studied previously [2]. This study has analyzed some additional ones. Of these systems $1 + D - D^2 - D^3$ deserves more attention since its performance rates above some of the previously studied PRS systems, for instance Kretzmer's class 5 system [2]. The two systems which still stand out on the basis of the performance measures used here are duobinary $1 + D$ and modified duobinary $1 - D^2$. These are also the ones that have in the past been put to practice both because of their simplicity and their useful spectral shapes.

The exact figures for probability of error and hence SNR degradation have been presented for one system

model. This allows us to evaluate the true effect of precoding. The results show that for reasonable error rates (10^{-5} or below for instance), precoding does not decrease the SNR degradation greatly and in some cases actually increases it. (This is due in part to the fact that error propagation affects the error rate by at most a multiplicative factor.) However, since precoding is relatively easy to implement, it serves a useful purpose for many PRS systems.

In real systems since the data are not truly random, selected input sequences could conceivably suffer severe error propagation. Precoding does offer protection against this and in addition tends to scramble some repeated data patterns, especially in the case of systems with the more complicated system polynomials.

Detectors such as the maximum likelihood sequence estimator were not emphasized here. Until the high cost of implementing these types of detectors decreases substantially. PRS systems will continue to use decoding schemes as described here. Presently, PRS systems still tend to be applied where simplicity of processing and detecting is important. However, in the future, we can look forward to achieving better performance for PRS with more optimal decoding schemes.

ACKNOWLEDGMENT

REFERENCES

[1] A. Lender, "The duobinary technique for high-speed data transmission," *IEEE Trans. Commun. Electron.*, vol. 82, pp. 214–218, May 1963.
[2] E. R. Kretzmer, "Generalization of a technique for binary data communication," *IEEE Trans. Commun. Technol.* (Concise Papers), vol. COM-14, pp. 67–68, Feb. 1966.
[3] T. L. Swartz, "Performance analysis of a three-level modified duobinary digital FM microwave radio system," in *Conf. Rec., IEEE Int. Conf. Communications*, 1974, pp. 5D-1–5D-4.
[4] F. K. Becker, E. R. Kretzmer, and J. R. Sheehan, "A new signal format for efficient data transmission," *Bell Syst. Tech. J.*, vol. 45, pp. 755–758, May–June 1966.
[5] J. W. Smith, "Error control in duobinary systems by means of null zone detection," *IEEE Trans. Commun. Technol.*, vol. COM-16, pp. 825–830, Dec. 1968.
[6] J. F. Gunn and J. A. Lombardi, "Error detection for partial-response systems," *IEEE Trans. Commun. Technol.*, vol. COM-17, pp. 734–737, Dec. 1969.
[7] H. Kobayashi and D. T. Tang, "On decoding of correlative level coding systems with ambiguity zone detection," *IEEE Trans. Commun. Technol.*, vol. COM-19, pp. 467–477, Aug. 1971.
[8] A. M. Gerrish and R. D. Howson, "Multilevel partial-response signaling," in *Conf. Rec., IEEE Int. Conf. Communications*, 1967, p. 186.
[9] M. Tomlinson, "New automatic equalizer employing modulo arithmetic," *Electron. Lett.*, vol. 7, nos. 5/6, pp. 138–139, Mar. 1971.
[10] H. Harashima and H. Miyakawa, "Matched-transmission technique for channels with intersymbol interference," *IEEE Trans. Commun.*, vol. COM-20, pp. 774–780, Aug. 1972.
[11] H. Kobayashi, "A survey of coding schemes for transmission or recording of digital data," *IEEE Trans. Commun. Technol.*, vol. COM-19, pp. 1087–1100, Dec. 1971.
[12] ——, "Correlative level coding and maximum-likelihood decoding," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 586–594, Sept. 1971.
[13] G. D. Forney, Jr., "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 363–378, May 1972.
[14] R. W. "Chang, A new equalizer structure for fast start-up digital communication," *Bell Syst. Tech. J.*, vol. 50, pp. 1969–2014, July–Aug. 1971.
[15] K. H. Mueller, "A new, fast-converging mean-square algorithm for adaptive equalizers with partial-response signaling," *Bell Syst. Tech. J.*, vol. 54, pp. 143–153, Jan. 1975.
[16] O. B. P. Rikkert de Koe and P. van der Wurf, "On some extensions of Nyquist's telegraph transmission theory," *Proc. IEEE* (Lett.), vol. 57, pp. 701–702, Apr. 1969.
[17] W. R. Bennett, *Introduction to Signal Transmission*. New York: McGraw-Hill, 1970, p. 16.
[18] D. M. Baker, "Analog/digital hybrid radio," *Can. Electron. Eng.*, pp. 34–38, Feb. 1973.
[19] D. A. George, R. R. Bowen, and J. R. Storey, "An adaptive decision feedback equalizer," *IEEE Trans. Commun. Technol.*, vol. COM-19, pp. 281–293, June 1971.
[20] D. L. Duttweiler, J. E. Mazo, and D. G. Messerschmitt, "An upper bound on the error probability in decision-feedback equalization," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 490–497, July 1974.
[21] W. Feller, *An Introduction to Probability Theory and Its Applications*, 2nd ed., vol. 1. New York: Wiley, 1957.
[22] G. Birkoff and S. MacLane, *A Survey of Modern Algebra*, 3rd ed. New York: Macmillan, 1965.
[23] P. Kabal and S. Pasupathy, "Partial-response signalling," Dep. Elec. Eng., Univ. of Toronto, Toronto, Ont., Canada, Commun. Tech. Rep. 75-1, Jan. 1975.
[24] R. W. Lucky, J. Salz, and E. J. Weldon, Jr., *Principles of Data Communication*. New York: McGraw-Hill, 1968.
[25] B. M. Smith, "Some results for the eye patterns of class 4 partial response data signals," *IEEE Trans. Commun.* (Concise Papers), vol. COM-22, pp. 696–698, May 1974.
[26] J. W. Craig, "Eye patterns and timing error sensitivity in partial-response pulse transmission systems," in *Conf. Rec., IEEE Int. Conf. Communications*, 1975.
[27] S. U. H. Qureshi and E. E. Newhall, "An adaptive receiver for data transmission over time-dispersive channels," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 448–457, July 1973.

★

Peter Kabal was born in Stockholm, Sweden, on November 3, 1945. He received the B.A.Sc. and M.A.Sc. degrees in electrical engineering from the University of Toronto, Toronto, Ont., Canada, in 1968 and 1972, respectively. He is currently completing the requirements for the Ph.D. degree at the University of Toronto. He has held National Research Council of Canada scholarships while at graduate school.

He worked for the Hydro-Electric Power Commission of Ontario from 1968 to 1969. He has also held a number of summer positions including ones at the Ontario Hydro Research Laboratories, Toronto, and at Bell Laboratories, Murray Hill, N.J. He has been a Teaching Assistant and a part-time Lecturer at the University of Toronto. His research interests are in the field of digital communication.

★

Subbarayan Pasupathy (M'73) was born in Madras, India, on September 21, 1940. He received the B.E. degree in telecommunications from the University of Madras, Madras, India, in 1963, the M. Tech. degree in electrical engineering from the Indian Institute of Technology, Madras, India, in 1966, and the M. Phil. and Ph.D. degrees in engineering and applied science from Yale University, New Haven, Conn., in 1970 and 1972, respectively.

During 1965–1967, he was a Research Scholar at the Indian Institute of Technology, and during 1968–1971 he was a Teaching Assistant at Yale University. During 1972–1973, he was a Postdoctoral Fellow at the University of Toronto, Toronto, Ont., Canada, where he is now an Assistant Professor of Electrical Engineering. His current research interests include statistical communication theory, array processing of signals, radar-sonar systems, and data communication.