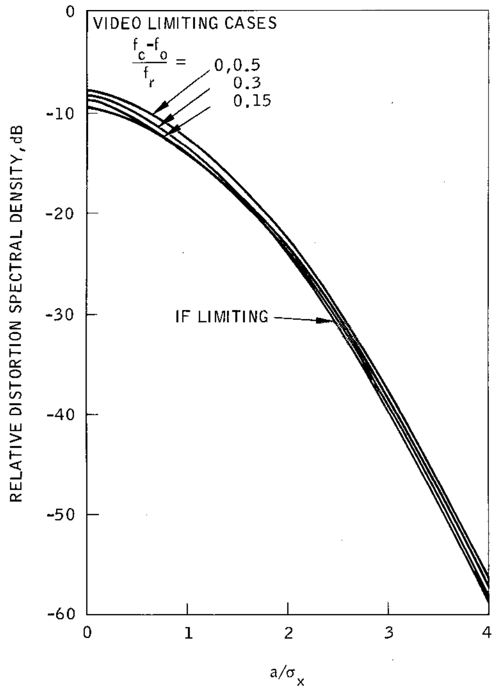

Fig. 13. Ratio of distortion to signal spectral densities at center of band, $B_x/f_r = 0.2$.

distortion spectral density outside the input spectral region, what in the radar application would be the MTI problem. Also of interest, for example in a mapping radar, is the ratio of signal-to-distortion spectral densities near the center of the signal spectrum, where signal refers to the random input signal. The distortion component in this region is shown dashed in the figures. Figs. 12 and 13 are cross plots of the relative level of distortion as a function of the limiter drive a/σ_x. In this case the video limiter is seen to produce only slightly higher distortion spectral levels. The results are relatively insensitive to input bandwidth B_x/f_r.

The results presented here indicate that a coherent two-channel signal processor, when designed to separate a small signal from a larger random signal, may perform better with an intentional IF limiter if there is significant A/D converter saturation.

References

[1] W. B. Davenport, Jr., "Signal-to-noise ratios in band-pass limiters," *J. Appl. Phys.*, vol. 24, June 1953, pp. 720–727.
[2] N. M. Blachman, "The output signal-to-noise ratio of a power-law device," *J. Appl. Phys.*, vol. 24, June 1953, pp. 783–785.
[3] ——, "The effect of a limiter upon signals in the presence of noise," *IRE Trans. Inform. Theory* (Corresp.), vol. IT-6, March 1960, p. 52.
[4] C. R. Cahn, "A note on signal-to-noise ratios in band-pass limiters," *IRE Trans. Inform. Theory*, vol. IT-7, Jan. 1961, pp. 39–43.
[5] W. Doyle, "Elementary derivation for band-pass limiter S/N," *IRE Trans. Inform. Theory* (Corresp.), vol. IT-8, Apr. 1962, p. 259.
[6] N. M. Blachman, "Band-pass nonlinearities," *IRE Trans. Inform. Theory*, vol. IT-10, Apr. 1964, pp. 162–164.
[7] D. Middleton, *An Introduction to Statistical Communication Theory*. New York: McGraw-Hill, 1960, pp. 574–594.
[8] W. B. Davenport and W. L. Root, *Random Signals and Noise*. New York: McGraw-Hill, 1958, chs. 8, 13.
[9] C. W. Helstrom, *Statistical Theory of Signal Detection*. New York: Pergamon, 1968, ch. 2, sect. 6.

Correlative Level Coding and Maximum-Likelihood Decoding

HISASHI KOBAYASHI, MEMBER, IEEE

***Abstract*—Modems for digital communication often adopt the so-called correlative level coding or the partial-response signaling, which attains a desired spectral shaping by introducing controlled intersymbol interference terms. In this paper, a correlative level encoder is treated as a linear finite-state machine and an application of the maximum-likelihood decoding (MLD) algorithm, which was originally proposed by Viterbi in decoding convolutional codes, is discussed. Asymptotic expressions for the probability of decoding error are obtained for a class of correlative level coding systems, and the results are confirmed by computer simulations. It is shown that a substantial performance gain is attainable by this probabilistic decoding method.**

Manuscript received August 6, 1970.
The author is with the IBM Thomas J. Watson Research Center, Yorktown Heights, N.Y. 10598.

I. Introduction

A TECHNIQUE in digital data communication developed in recent years is the so-called correlative level coding (Lender [1]) or the partial-response channel signaling (Kretzmer [2]). This signaling method is different from the conventional pulse-amplitude modulation (PAM) system in that a controlled amount of intersymbol interference is introduced to attain a certain beneficial spectral shaping. Such a system possesses in general the property of being relatively insensitive to channel imperfections and to variations in transmission rate [3], [4]. Recently it has been pointed out [5] that a digital magnetic recording channel can be regarded also as a partial-response channel due to its inherent differentiation in the readback process.

Although the correlative level coding permits the transmission of data at the Nyquist rate (i.e., 2 Bd per cycle of bandwidth) or even at a higher rate in a practically band-limited channel, the increase in the number of signal levels results in loss of noise margin compared with binary antipodal signaling [1]–[3]. In any correlative level coding system, however, the coded output contains redundancy that can be utilized as a measure of error control at the receiving end. Lender [1] and Gunn and Lombard [6] discuss error detection methods for some special cases. Smith [7] has introduced the null-zone detection method in the duobinary system, in which most of the unreliable bits in the null zones are replaceable using the inherent redundancy of the correlative level coded sequence. A unified method for algebraic error control has been developed by Kobayashi and Tang [8], [9].

Recently an analogy between correlative level coding and convolutional coding has been pointed out by the present author [8], [10] and by Forney [11]. A correlative level encoder can be viewed as a simple type of linear finite-state machine defined over the real-number field as opposed to a Galois field over which a convolutional encoder is defined. The present paper will show that the maximum-likelihood decoding (MLD) algorithm devised by Viterbi [12], [13] in decoding convolutional codes is applicable to our problem. Both analytical and experimental results of this probabilistic decoding scheme will be presented. The performance of the maximum-likelihood decoding is much superior to any other method reported thus far. Asymptotic expressions for the decoding error probability are derived. Several other important problems associated with the MLD method are discussed: the effect of precoding on the decoding error rate and error patterns, the number of quantization levels required, and the problem of decoder buffer overflows.

II. Mathematical Model of Correlative Level Coding Systems [5]

A sequence is represented by a power series in Huffman's delay operator D. An information sequence $\{a_k\}$ is then represented by

$$A(D) = \sum_{k=1}^{\infty} a_k D^k. \tag{1}$$

A correlative level encoder or a partial-response channel is a linear system that is characterized by a transfer function

$$G(D) = \sum_{i=0}^{L} g_i D^i, \tag{2}$$

where g_i are integers with a greatest common divisor equal to one. Given an input sequence $A(D)$, the encoded output sequence $X(D) = \sum_{k=1}^{\infty} x_k D^k$ is determined by

$$X(D) = G(X) \cdot A(D). \tag{3}$$

Among $(L + 1)$ coefficients of $G(D)$ the leading coefficient g_0 represents the signal value while other L coefficients correspond to controlled intersymbol interference terms. Let the size of the source alphabet be m. One can choose a set of integers $\{0,1,\cdots, m-1\}$ as the alphabet without loss of generality. Then the signal levels that the encoder output $X(D)$ takes on range from $(m-1)\sum_{i=0}^{L} \min\{g_i,0\}$ to $(m-1)\sum_{i=0}^{L} \max\{g_i,0\}$, that is, the size of the channel alphabet is $M = (m-1)\sum_{i=1}^{L} |g_i| + 1$.

The encoder output $X(D)$ is sent over a channel with an additive noise sequence $Z(D)$ (Fig. 1), the output of which is denoted by $Y(D)$:

$$Y(D) = X(D) + Z(D). \tag{4}$$

At the receiving end the sequence $X(D)$ is first led to a quantizer whose output is denoted by $Q(D)$. Let us consider the case in which a quantizer makes a hard decision, i.e., it assigns to each y_k one of M possible integers. If no errors are introduced in the channel and quantizer, then $Q(D) = X(D)$ and thus the information sequence $A(D)$ can be recovered simply by passing $Q(D)$ through the inverse filter $1/G(D)$. An immediately apparent drawback of the inverse filtering is that if $Q(D)$ contains an error the effect of this error tends to propagate in the decoded sequence. This is easily seen from the fact that $1/G(D)$ can be expanded, in general, into an infinite power series in D. A scheme for avoiding this error propagation is the so-called "precoding" devised by Lender [1] originally for the duobinary system, i.e., $G(D) = 1 + D$ and $m = 2$.

The precoding operation can best be described in terms of a discrete filter with transfer function $[1/G(D)]_{\mathrm{mod}\ m}$, which is nonlinear in the ordinary sense but is linear over the residue class ring modulo m. In order that this filter exist, it is required that the inverge of g_0 exists in the residue class ring modulo m. This is assured if g_0 and m are relatively prime. With an m-level input sequence $A(D)$, the precoded sequence $B(D)$, which is also m level, is determined by

$$B(D) = A(D)/G(D), \qquad \mathrm{mod}\ m \tag{5a}$$

or equivalently

$$G(D)B(D) = A(D), \qquad \mathrm{mod}\ m. \tag{5b}$$

The correlative level encoder transforms $B(D)$ into $X(D)$ according to the relation

$$G(D)B(D) = X(D). \tag{6}$$

It immediately follows from (5) and (6) that

$$X(D) = A(D), \qquad \mathrm{mod}\ m \tag{7a}$$

or

$$x_k = a_k, \qquad \mathrm{mod}\ m, \text{ for all } k. \tag{7b}$$

To recover the information sequence $A(D)$ given the hard decision output $Q(D)$, one merely performs "mod m" operation on $Q(D)$ the output of which is denoted by $\hat{A}(D)$ (Fig. 2). Propagation of errors in the output $\hat{A}(D)$ is thus eliminated.

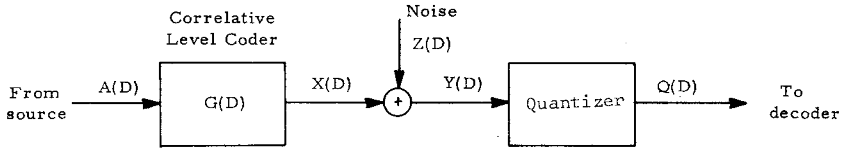

Fig. 1. A discrete system representation of a correlative level coding or partial-response signaling system.

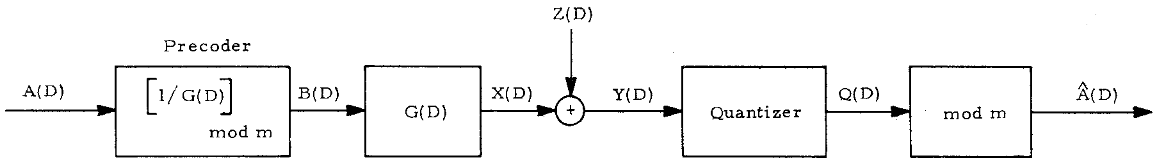

Fig. 2. Correlative level coding system with a precoder and "mod m" detector (the conventional bit-by-bit detection method configuration).

An algebraic method of error detection [5], [8] makes full use of the inherent redundancy of an M-level sequence $X(D)$. This algebraic approach has been further extended to the case in which the receiver makes a soft decision, including ambiguity levels [8], [9].

The present paper describes a completely different approach to decoding correlative level coded sequences, namely, the MLD method based on a linear finite-state machine representation of a correlative level encoder.

III. Maximum-Likelihood Decoding Algorithm

In 1967 Viterbi [12] devised a new nonsequential decoding algorithm for convolutional codes. Forney [14] showed that this algorithm is in fact the MLD rule. Omura [15] discussed the algorithm in a state–space context and showed its equivalence to dynamic programming.

A correlative level encoder defined in the previous section can be regarded as a linear finite-state machine like a convolutional encoder. In convolutional codes redundancy is introduced timewise, whereas in correlative level coding redundancy is induced amplitudewise. For a given correlative level encoder $G(D)$ of (2), we define s_k the "state" of the encoder by the latest L input digits, i.e.,

$$s_k = [a_{k-L+1}, a_{k-L+2}, \cdots, a_k] \tag{8a}$$

or

$$s_k = [b_{k-L+1}, b_{k-L+2}, \cdots, b_k] \tag{8b}$$

depending on whether a precoder is included or not. Then the encoder output x_k is a function of s_{k-1} and a_k.[1] In the following we limit ourselves to the correlative level coding system $G(D) = 1 - D$. This transfer function corresponds to a digital magnetic recording channel [5], [10]. The so-called modified duobinary [1] or the partial-response channel [2] class IV corresponds to the transfer function $G(D) = 1 - D^2$, which is merely an interleaved form of $1 - D$. The duobinary signaling [1] $1 + D$ holds a dual relationship with $1 - D$ as is shown in Appendix I. Thus the basic decoder structure and the performance are common to the class $G(D) = 1 \pm D^K$, $k = 1,2,\cdots$.

Now the definition of the state s_k is reduced to

$$s_k = a_k \qquad \text{(without precoding)} \tag{9a}$$

or

$$s_k = b_k = s_{k-1} + a_k, \qquad \text{mod } m \text{ (with precoding)}. \tag{9b}$$

Fig. 3 shows the transition of the encoder state as a function of time k. Starting from $s_0 = 0$, the encoder follows a particular path according to the input sequence $A(D)$. Note that the channel sequence $X(D)$ is given by the *slope* of branches in the trellis picture [14] since

$$X(D) = (1 - D)S(D), \text{ i.e., } x_k = s_k - s_{k-1}. \tag{10}$$

Let us consider an input sequence of length $N, A^N(D) = \sum_{k=1}^{N} a_k D^k$. If each digit can take on $0,1,\cdots, m-1$ with equal probability and independently, there are m^N different paths on the trellis picture of Fig. 3. An optimum decoder will be the one that chooses the most likely path based on the channel output sequence $Y^N(D)$. Throughout the following discussion we assume that noise is independent from digit to digit. Then the log-likelihood function of a path $S^N(D) = \sum_{k=0}^{N} s_k D^k$ is given as the sum of the log-likelihood functions of N transitions

$$l(Y^N(D) \mid S^N(D)) = \sum_{k=0}^{N} l(y_k \mid s_{k-1}, s_k). \tag{11}$$

The MLD algorithm proceeds as follows. Starting from the known initial state s_0, the decoder computes $l(y_1 \mid s_0, s_1)$ for $s_1 = 0,1,\cdots, m-1$. We define the metric of the node $s_1 = i$ by

$$\mu_1(i) = l(y_1 \mid s_0, i), \qquad i = 0,1,\cdots, m-1. \tag{12}$$

In general, at time k (≥ 2), the decoder compares for each node $s_k = j$ the log-likelihood functions of m different paths leading to $s_k = j$, i.e., $\mu_{k-1}(i) + l(y_k \mid i,j)$, $i = 0,1,\cdots, m-1$. Let the path with the largest likelihood

[1] In the literature of finite-state machines [16] either the Moore machine model or the Mealy machine model is usually adopted. In the former model we define the state by $s_k = [a_{k-L}a_{k-L+1}\cdots a_k]$, thus x_k is a function of s_k only. In the latter case we define $s_k = [a_{k-L}a_{k-L+1}\cdots a_{k-1}]$, thus x_k is a function of s_k and a_k. We find the definition of (8) most convenient in our problem.

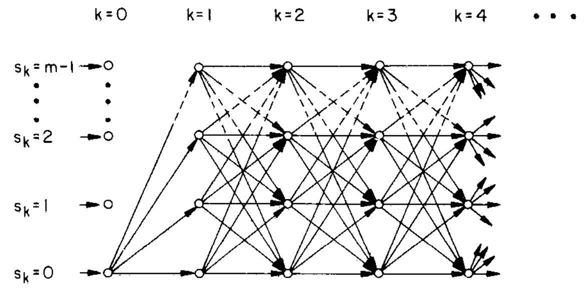

Fig. 3. Trellis picture representation of system $G(D) = 1 - D$ with an m-level input: $x_k = s_k - s_{k-1}$ where $s_k = a_k$ for a system without precoding, and $s_k = b_k = a_k + s_{k-1}$ (mod m) for a system with precoding.

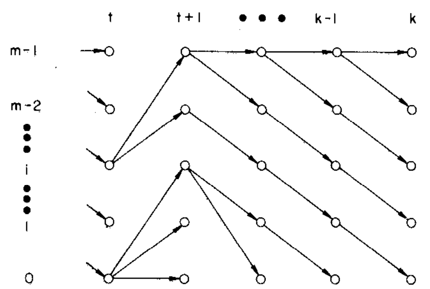

Fig. 4. Illustration of an event in which all surviving paths at time k have the common node at $s_t = i$ [see (14)].

function be called the survivor [12], since only this branch has the possibility of being a portion of the maximum-likelihood path and hence should be preserved. Other $(m - 1)$ paths ending at $s_k = j$ can be discarded. Thus the metric of the survivor at $s_k = j$ is

$$\mu_k(j) = \max_i \{\mu_{k-1}(i) + l(y_k \mid i,j)\}, \tag{13}$$

where $j = 0,1,\cdots, m - 1$ and $k \geq 2$. Each time an event occurs such that all surviving paths branch out from a common node, say $s_t = i$, that is, the following relation

$$\mu_k(j) = \mu_t(i) + l(y_{t+1},\cdots,y_{k-1}y_k \mid i,s_{t+1}^j \cdots s_{k-1}^j,j) \tag{14}$$

holds at time k for all $j = 0,1,\cdots, m - 1$ (see Fig. 4), then the maximum-likelihood path up to time t is uniquely determined independently of the succeeding digits, where $[i,s_{t+1}^j \cdots s_{k-1}^j,j]$ is the surviving path ending at $s_k = j$.

Now we are in a position to discuss a practical implementation of the maximum-likelihood decoder. Assume that the additive noise $Z(D)$ of the channel is a stationary Gaussian random sequence with zero mean and variance σ^2. Let the signal level spacing in the channel be A instead of unity. Then the log-likelihood function is simplified to

$$\begin{aligned} l(y_k \mid s_{k-1},s_k) &= \ln p_z(y_k - A(s_k - s_{k-1})) \\ &= -\frac{1}{2\sigma^2}\{y_k - A(s_k - s_{k-1})\}^2 - \tfrac{1}{2}\ln(2\pi\sigma^2). \end{aligned} \tag{15}$$

Notice that the terms $-(1/2\sigma^2)y_k^2 - \frac{1}{2}\ln(2\pi\sigma^2)$ are common to the log-likelihood function of all the branches and

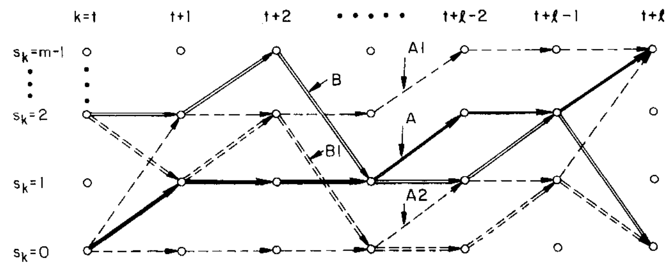

Fig. 5. Correct paths A and B and their adversary paths $A1$, $A2$, and $B1$. Path A takes neither state 0 nor $m - 1$ in the interval $t + 1 \leq k \leq t + l - 1$. Path B does not take state $m - 1$ in the same interval.

hence can be deleted. Furthermore, by dividing the resultant terms by a constant $A/2\sigma^2$, we have a simplified version of the recursive formula for metrics

$$\tilde{\mu}_k(j) = \max_i \{\tilde{\mu}_{k-1}(i) + (j - i)y_k - \tfrac{1}{2}(j - 1)^2 A\},$$
$$j = 0,1,\cdots, m - 1 \text{ and } k \geq 1 \tag{16}$$

and

$$\tilde{\mu}_0(j) = \begin{cases} 0, & j = s_0 \\ -\infty, & j \neq s_0. \end{cases} \tag{17}$$

The structure of the maximum-likelihood decoder for a binary input is discussed in detail in [10].

IV. Performance of Maximum-Likelihood Decoder

In the present section we present analytical results on the performance[2] of the MLD algorithm and then computer simulation results will be reported to confirm the analytical results. We see from the trellis picture of Fig. 3 that in the MLD method an error occurs when and only when the decoder path diverges from the correct path at some time, say $k = t$. They remerge at some time later at $k = t + l$, $l \geq 2$.

Assume a high signal-to-noise ratio (SNR) condition. Then consider, as possible *adversary* paths (paths competing against the correct path) only those that are "closest" to the correct path. Since the slope of the trellis corresponds to the signal level in the channel, adversaries are those that stay closest to, and parallel with, the correct path in the interval $t + 1 \leq k \leq t + l - 1$.

There are at most two adversaries for a given correct path. Let the path A of Fig. 5 be the correct path. Note that this path does not take on extreme states 0 and $m - 1$ between $k = t + 1$ and $t + l - 1$. Then there are exactly two adversary paths $A1$ and $A2$ that diverge from the correct path A at $k = t$ and remerge for the first time at $k = t + l$. Let the metric of an adversary path $[s_t\tilde{s}_{t+1} \cdots \tilde{s}_{t+l-1}s_{t+l}]$ minus that of the correct path $[s_ts_{t+1} \cdots s_{t+l-1}s_{t+l}]$ be denoted by w_l, then w_l is a random variable given, from (15), by

$$w_l = v_1 + v_l, \tag{18}$$

[2] The author has found through a private communication that the same results have been recently obtained by Forney [17].

where

$$v_1 = (\tilde{s}_{t+1} - s_{t+1})y_{t+1} - \tfrac{1}{2}A\{(\tilde{s}_{t+1} - s_t)^2 - (s_{t+1} - s_t)^2\} \quad (19)$$

and

$$v_l = (s_{t+l-1} - \tilde{s}_{t+l-1})y_{t+l} - \tfrac{1}{2}A\{(s_{t+l} - \tilde{s}_{t+l-1})^2 - (s_{t+l} - s_{t+l-1})^2\}. \quad (20)$$

Equation (18) is obtained using the fact that x_k for $t + 2 \leq k \leq t + l - 1$ are common to both the correct path and the adversary, hence these branches do not contribute to the quantity w_l. Random variables v_1 and v_l are Gaussian with

$$E\{v_1\} = E\{v_l\} = -\tfrac{1}{2}A \quad (21)$$

and

$$\operatorname{var}\{v_1\} = \operatorname{var}\{v_l\} = \sigma^2. \quad (22)$$

Hence w_l is a Gaussian random variable with

$$E\{w_l\} = -A \quad (23)$$

and

$$\operatorname{var}\{w_l\} = 2\sigma^2. \quad (24)$$

The adversary path $[s_t\tilde{s}_{t+1} \cdots \tilde{s}_{t+l-1}s_{t+l}]$ wins over the correct path if $w_l > 0$, and such an event occurs with probability

$$P_e(l) = \int_0^\infty (1/\sqrt{2}\sigma)\phi((w_l + A)/(2\sigma^2)^{1/2})\, dw_l = Q(d), \quad (25)$$

where $\phi(\cdot)$ is the unit normal function and $Q(\cdot)$ is the function defined by

$$Q(\chi) = \int_\chi^\infty \phi(t)\, dt. \quad (26)$$

The quantity d of (25) is the SNR parameter defined by

$$d^2 = A^2/(2\sigma^2). \quad (27)$$

If no precoding is used, $(l - 1)$ consecutive digits $a_{t+1}a_{t+2} \cdots a_{t+l-1}$ are erroneously decoded when either of two adversaries *A1* and *A2* wins over the correct path *A*. A union bound for the average number of error digits due to such events is given by

$$P_1 = 2 \sum_{l=2}^{\infty} (l - 1) \left(\frac{m-2}{m}\right)^{l-1} P_e(l) = \tfrac{1}{2}m(m - 2)Q(d) \quad (28)$$

in which $[(m - 2)/m]^{l-1}$ is the probability that the correct path *A* does not take extreme values 0 and $m - 1$. If precoding is adopted in the system, a_{t+1} and a_{t+l} are erroneously decoded when either *A1* or *A2* wins over *A*. Such a probability is bounded by

$$P_1' = 2 \sum_{l=2}^{\infty} 2 \left(\frac{m-2}{m}\right)^{l-1} P_e(l) = 2(m - 2)Q(d). \quad (29)$$

If a correct path takes one of two outermost states 0 and $m - 1$ but never takes both between $k = t + 1$ and $t + l - 1$, then an adversary to be considered is only one. For example, if the path *B* is the correct one, then we need consider only *B1*. Decoding error rate due to such an event is given by

$$P_2 = \sum_{l=2}^{\infty} (l - 1) \left\{2\left(\frac{m-1}{m}\right)^{l-1} - 2\left(\frac{m-2}{m}\right)^{l-1}\right\} P_e(l) = 2m(m - 1)Q(d) - P_1 \quad (30)$$

when precoding is not used and

$$P_2' = \sum_{l=2}^{\infty} 2 \left\{2\left(\frac{m-1}{m}\right)^{l-1} - 2\left(\frac{m-2}{m}\right)^{l-1}\right\} P_e(l) = 4(m - 1)Q(d) - P_1' \quad (31)$$

when precoding is adopted. The term $\{2[(m - 1)/m]^{l-1} - 2[(m - 2)/m]^{l-1}\}$ in (30) and (31) represents the probability that the correct path *B* takes one of two outermost states 0 and $m - 1$ at least once, but never takes the other extreme state, in the interval $l + 1 \leq k \leq t + l - 1$.

Adding (28) and (30), the symbol error rate is given by

$$P_{\mathrm{MLD}} = 2m(m - 1)Q(d) \quad (32)$$

for a system without precoding. Similarly from (29) and (31) the symbol error probability of a system with precoding is

$$P'_{\mathrm{MLD}} = 4(m - 1)Q(d). \quad (33)$$

Except for $m = 2$ (i.e., binary inputs), $P'_{\mathrm{MLD}} < P_{\mathrm{MLD}}$. Thus precoding is beneficial not only in the conventional bit-by-bit detection method but also in the MLD method.

In the bit-by-bit detection method the error rate is given by [3]

$$P_{\mathrm{BIT}} = 2(1 - 1/m^2)Q(d/2^{1/2}). \quad (34)$$

The average SNR R in the channel is given by

$$R = [(m^2 - 1)/3]\, d^2. \quad (35)$$

Therefore from (33) to (35) expressions for decoding error rates in terms of the channel SNR are given by

$$P'_{\mathrm{MLD}} = 4(m - 1)Q([3R/(m^2 - 1)]^{1/2}) \quad (36)$$

and

$$P_{\mathrm{BIT}} = 2(1 - 1/m^2)Q([3R/2(m^2 - 1)]^{1/2}). \quad (37)$$

It will be interesting to compare these results with an m-level PAM system without correlative level coding. The expression for the error rate is [3]

$$P_m = 2(1 - 1/m)Q([3R/(m^2 - 1)]^{1/2}). \quad (38)$$

When $m = 2$, P'_{MLD} is only four times of P_m; thus the MLD method allows a PAM system to adopt a correlative level coding technique to attain some desired spectral shaping with a very little penalty in its performance. In other words, the loss in noise margin, which has been claimed to be the major disadvantage of a correlative level coding system, can be almost completely recovered by the MLD method.

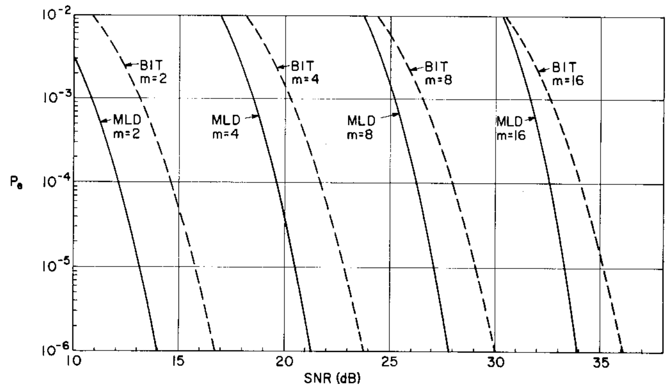

Fig. 6. Probability of error versus SNR for the bit-by-bit detection method and the MLD method in a system $G(D) = 1 \pm D^K$ with an m-level input, where $m = 2$, 4, 8, and 16.

Fig. 6 plots (36) and (37) for $m = 2$, 4, 8, and 16. As we see from this figure, the MLD improves the performance significantly compared with the bit-by-bit detection method. For a binary input, for example, $P_{\mathrm{BIT}} = 1.1 \times 10^{-3}$ at $R = 13$ dB, whereas $P'_{\mathrm{MLD}} = 1.8 \times 10^{-5}$, i.e., the performance improvement by a factor of 70. This factor becomes as high as 250 for $R = 14$ dB.

Computer simulations were done for a binary input, and the results are plotted in Fig. 7. The simulation size was $N = 10^5$ for $R = 10 - 12$ dB and $N = 10^6$ for R greater than 12 dB. Although this sample size is not sufficient, we may conclude that the analysis and the experiment agree satisfactorily. In the above simulation the decoder buffer length was 25 to avoid a possible buffer overflow. The problem of buffer overflows will be discussed in the next section.

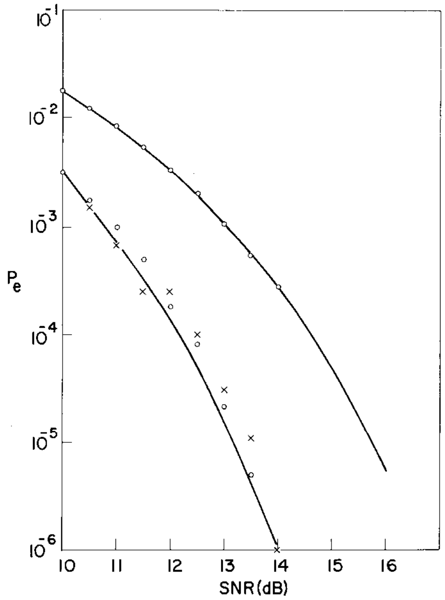

Fig. 7. Experimental curves of the bit-by-bit detection method and the MLD method for a system $G(D) = 1 - D$ with $m = 2$. x—no precoding, o—precoding.

V. Some Practical Considerations

In the present section we will consider two important questions that would invariably arise when one wants to implement the maximum-likelihood decoder. The first problem is the number of quantization levels required. Thus far we have tacitly assumed that the receiver input y_k is quantized into infinitely many levels. In an actual implementation, which is presumably in a digital form, the channel output y_k must be quantized into a finite number of levels. If the quantizer output is denoted by q_k (see Fig. 1), then a discrete memoryless channel (DMC) can be defined with input x_k and output q_k. The MLD rule applied to the output of this DMC is given by [see (13)]

$$\mu_k(j) = \max_i \{\mu_{k-1}(i) + \ln p[q_k \mid x_k = (j - i)]\} \quad (39)$$

where $p(\cdot \mid \cdot)$ is the channel transition probability of the DMC defined above. The decoding rule (39) is applicable regardless of the number of quantization levels, and whether or not the quantization level spacing is uniform. However, the metric computation may not be practical in this form, since it needs the table of $\ln p(\cdot \mid \cdot)$ and these numbers require several significant figures to represent.

A more practical scheme will be the one that uses the simple formula of (16) (y_k is to be replaced by q_k). If a uniform quantization is performed, the metric computation of (16) can be done in the integer format. Fig. 8 shows the performance of the MLD for a system with $G(D) = 1 \pm D^K$ and $m = 2$, when the quantization spacing is A/N where $N = 4$, 8, 16, and 32. We may conclude that $N = 16$ achieves almost the same performance (less than 0.1-dB loss) as the infinite level quantization.

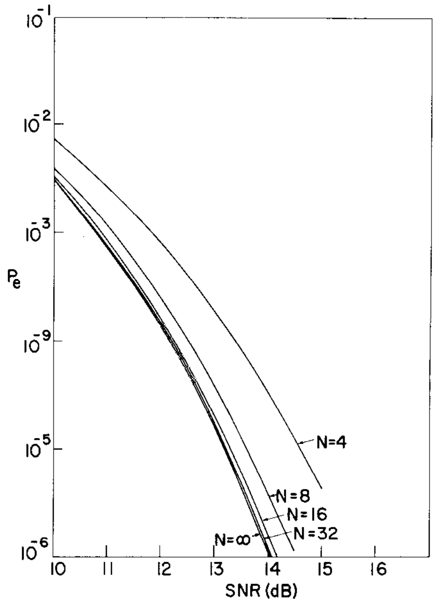

Fig. 8. Performance of the MLD decoder for a system with a finite number of quantization levels. Quantization level spacing is A/N where $N = 4$, 8, 16, and 32, and A is the signal level spacing.

The second problem that would arise in implementation will be the memory size required in the decoder. The decoder can store only a finite history, say the past J digits, of m surviving paths. A buffer overflow occurs when and only when events defined by (14) are separated by more

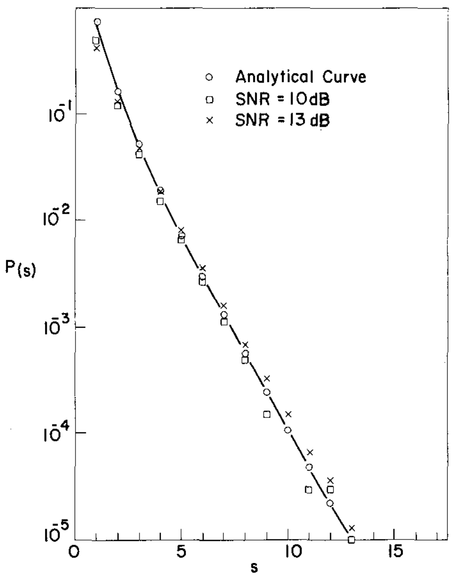

Fig. 9. Plot of (40) and experimental results for SNR = 10 dB and 13 dB.

than J time units. Let t and t' ($t' \geq t$) be two consecutive times at which m survivors branch out of a common node. Then for a binary input system (i.e., $m = 2$) the distribution of separation $s = t' - t$ is given by

$$P(s) = 2^{-s}\left[\frac{2}{s} - \frac{1}{s+1}\right] \cong \frac{1}{s}\,2^{-s}. \tag{40}$$

The derivation of (40) is given in Appendix II. Fig. 9 shows simulation results for SNR $R = 10$ and 13 dB along with the analytical curve of (40). We see a satisfactory agreement here also.

VI. Conclusions

The MLD of correlative level codes was discussed based on a finite-state machine representation of the encoder. A simple decoder structure was derived under a Gaussian noise assumption. Expressions (32) and (33) for the decoding error rate were obtained for a class of $G(D) = 1 \pm D$. It has been proven that a significant improvement in the performance is possible by the MLD method. Several important problems associated with the implementation of the decoder were also discussed; these include the effect of precoding, the number of quantization levels required, and the problem of decoder buffer overflows.

Appendix I
Duality of Systems $1 + D$ and $1 - D$

In actual systems the signal levels for $X(D)$ should be symmetrical around zero (dc) level to gain the largest separation under a given signal power constraint. Consider the mapping

$$c_k = b_k - \tfrac{1}{2}(m - 1), \tag{41}$$

which yields a dc-free sequence $\{c_k\}$. The performance of the system shown in Fig. 10(a) will be subject to no change when an alternating sequence $(-1)^k$ is multiplied at both

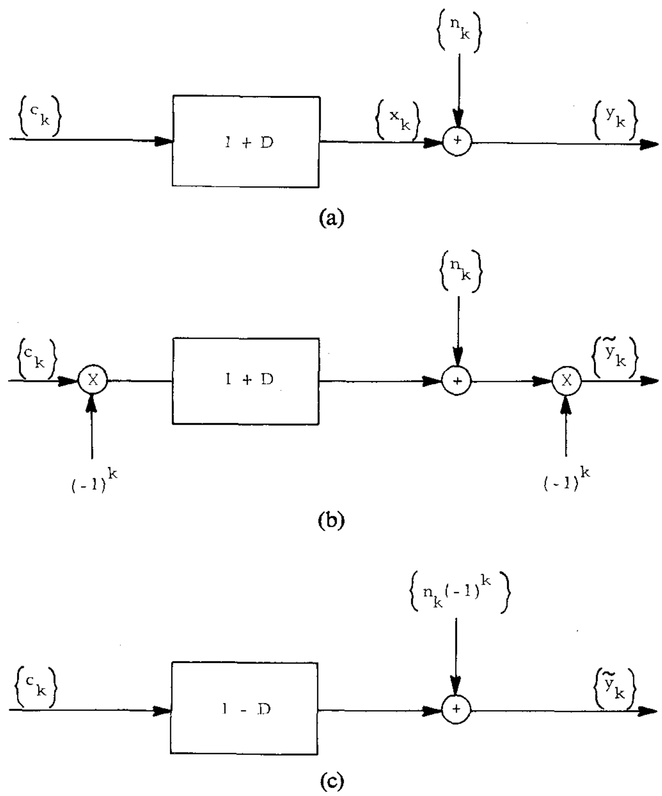

Fig. 10. (a) System $G(D) = 1 - D$ with a dc-free input. (b) System (a) preceded and followed by the modulation of a known sequence $(-1)^k$. (c) A system equivalent to (b).

the input and output. It is clear then that the system of Fig. 10(a) is equivalent to the system of Fig. 10(b), which is then simplified to Fig. 10(c). Thus if the noise sequence $\{n_k\}$ is uncorrelated, the system $1 + D$ will give the same performance with the system $1 - D$.

Appendix II
Derivation of (40)

Consider the binary input case. Referring to (14) and Fig. 4, we see that the earliest possible time k is always equal to $t + 1$, if a node i (either 0 or 1) at time t satisfies (14). This is because there are only two survivors for the binary case. Thus (14) is simplified to

$$\mu_{t+1}(j) = \mu_t(i) + l(y_{t+1} \mid i,j), \tag{42}$$

where $j = 0,1$ and $i = s_t$ is the common state from which two survivors of time $t + 1$ branch out [Fig. 11(a)]. Such an event occurs almost certainly when the correct path changes its state at $t + 1$, i.e., $s_{t+1} = 1 - s_t$. Fig. 11(a) shows the case $s_t = 1$ and $s_{t+1} = 0$. The reason for this is as follows. The correct path is almost always one of two survivors and the other survivor is very unlikely to grow from $s_t = 0$, since the transition from 0 to 1 corresponds to the channel symbol $+A$, which is the other extreme of the correct signal level $-A$. A necessary condition that a branching from a common node takes place at time $t' + 1$ for the first time, is that the correct path maintains its state at least up to time t', i.e., $s_k = s_{t+1}$ for $t + 2 \leq k \leq t'$. Such a path occurs with probability $2^{-(t'-t-1)} =$

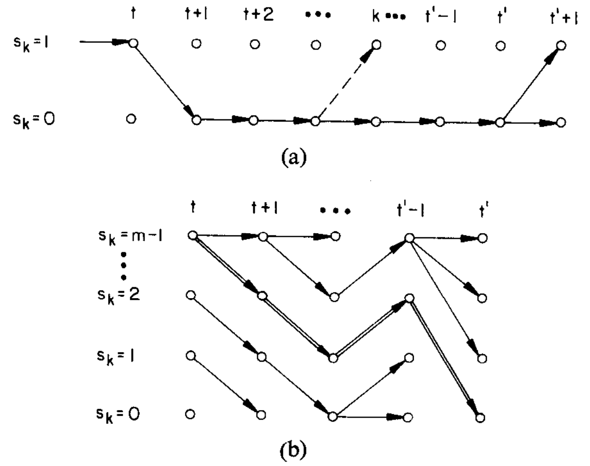

Fig. 11. (a) A correct path in a system $G(D) = 1 - D$ with a binary input. Branching of two survivors takes place at time t and t'. (b) Surviving paths in a system $G(D) = 1 - D$ with an m-level input. Path in a thick line represents a correct path.

2^{1-s}, where $s = t' - t$. Furthermore, the following condition must hold in order that a branching does not occur at $t + 2 \leq k \leq t'$:

$$(-y_{t+1} - \tfrac{1}{2}A) + (y_k - \tfrac{1}{2}A) < 0 \tag{43}$$

for $k = t + 2,\cdots,t'$.

Now at time $t' + 1$ the correct path moves to state 0 or state 1 with equal probability. If $s_{t'+1} = 1$ (i.e., a transition occurs), then a branching occurs almost surely as explained above. If $s_{t'+1} = 0$, then a branching occurs if and only if

$$(-y_{t+1} - \tfrac{1}{2}A) + (y_{t'+1} - \tfrac{1}{2}A) > 0. \tag{44}$$

Thus

$$\begin{aligned}\Pr\{&\text{the next branching occurs at } t'+1 \mid \text{a branching occurred at } t\} \\ &= 2^{1-s}\cdot 2^{-1}\Pr\{y_k - y_{t+1} < A, \quad k = t+2,\cdots,t'\} \\ &\quad + 2^{1-s}\cdot 2^{-1}\Pr\{y_k - y_{t+1} < A, \quad k = t+2,\cdots,t' \text{ and } y_{t'+1} - y_{t+1} > A\}. \end{aligned} \tag{45}$$

Random variables y_{t+1}, $y_k(t + 2 \leq k \leq t')$ and $y_{t'+1}$ are Gaussian with common variance σ^2 and mean $-A$, 0 and $+A$, respectively. Thus

$$\begin{aligned}\Pr\{y_k - y_{t+1} < A \text{ for all } k = t+2,\cdots,t'\} &= \int_{-\infty}^{\infty}\left(\prod_{k=t+2}^{t'}\Pr\{y_k \leq y_{t+1} + A\}\right)\frac{1}{(2\Pi\sigma^2)^{1/2}} \times \exp\left\{-\frac{(y_{t+1}+A)^2}{2\sigma^2}\right\}dy_{t+1} \\ &= \int_{-\infty}^{\infty}\Phi^{s-1}((y+A)/\sigma)\frac{1}{\sigma}\phi((y+A)/\sigma)\,dy = \frac{1}{s}, \end{aligned} \tag{46}$$

where $\phi(\cdot)$ is the unit normal function and $\Phi(\cdot)$ is

$$\Phi(\chi) = \int_{-\infty}^{\chi}\phi(t)\,dt.$$

Similarly

$$\begin{aligned}\Pr\{y_k - y_{k+1} < A, k = t+2,\cdots,t' \text{ and } y_{t'+1} - y_{t+1} > A\} &= \int_{-\infty}^{\infty}\Phi^{s-1}((y+A)/\sigma)\Phi(-(y+A)/\sigma)\frac{1}{\sigma} \times \phi((y+A)/\sigma)\,dy = \frac{1}{s} - \frac{1}{s+1}. \end{aligned} \tag{47}$$

Thus on substituting (46) and (47) into (45), the distribution of $s = t' - t$ is

$$P(s) = [2/s - 1/(s+1)]2^{-s} \tag{48}$$

and clearly

$$\sum_{s=1}^{\infty} P(s) = 1. \tag{49}$$

For a multilevel input, i.e., $m > 2$, the computation is rather complicated. An approximate computation of $P(s)$, however, is obtained as follows. Suppose the correct path is located at state node 0 (or $m - 1$) at time t. Then under a high SNR condition by the time the correct path reaches the other extreme state $m - 1$ (or 0), all the survivors are the ones that branch out from the common node s_t (see Fig. 11(b)). Therefore the branching node is almost always 0 and $m - 1$ of the correct path. Hence the distribution of separation s is approximately

$$P(s) = [(m-1)/m]^{s-1}\cdot 1/m. \tag{50}$$

Acknowledgment

The author wishes to thank Prof. A. J. Viterbi and Prof. J. K. Omura at the University of California, Los Angeles, for helpful discussions on the MLD algorithm. He is also indebted to Dr. G. D. Forney, Codex Corporation, Dr. D. T. Tang, IBM Research Center, and the reviewers for their constructive comments.

References

[1] A. Lender, "Correlative level coding for binary-data transmission," *IEEE Spectrum*, vol. 3, Feb. 1966, pp. 104–115.
[2] E. R. Kretzmer, "Generalization of a technique for binary-data transmission," *IEEE Trans. Commun. Technol.* (Concise Papers), vol. COM-14, Feb. 1966, pp. 67–70.
[3] R. W. Lucky, J. Salz, and E. J. Weldon, Jr., *Principles of Data Communication*. New York: McGraw-Hill, 1968, pp. 83–90.
[4] H. Kobayashi, "Coding schemes for reduction of intersymbol interference in data transmission systems," *IBM J. Res. Dev.*, vol. 14, July 1970, pp. 343–353.
[5] H. Kobayashi and D. T. Tang, "Application of partial-response channel coding to magnetic recording systems," *IBM J. Res. Dev.*, vol. 14, July 1970, pp. 368–375.
[6] J. F. Gunn and J. A. Lombardi, "Error detection for partial-response systems," *IEEE Trans. Commun. Technol.*, vol. COM-17, Dec. 1969, pp. 734–737.
[7] J. W. Smith, "Error control in duobinary data systems by means of null zone detection," *IEEE Trans. Commun. Technol.*, vol. COM-16, Dec. 1968, pp. 825–830.
[8] H. Kobayashi and D. T. Tang, "On decoding and error control of correlative level coding," presented at the 1970 Int. Symp. Information Theory, Noordwijk, the Netherlands, June 15–19, 1970; also IBM Thomas J. Watson Res. Ctr., Yorktown Heights, N.Y., Res. Rep. RC-3149, Nov. 1970.
[9] ——, "On decoding of correlative level coding systems with ambiguity zone detection," *IEEE Trans. Commun. Technol.*, vol. COM-19, Aug. 1971, pp. 467–477.

[10] H. Kobayashi, "Application of probabilistic decoding to digital magnetic recording systems," IBM Thomas J. Watson Res. Ctr., Yorktown Heights, N.Y., Res. Rep. RC-2986, Aug. 1970; also *IBM J. Res. Dev.*, vol. 15, Jan. 1971, pp. 64–74.
[11] G. D. Forney, Jr., "Error correction for partial response modems," presented at the 1970 Int. Symp. Information Theory, Noordwijk, the Netherlands, June 15–19, 1970.
[12] A. J. Viterbi, "Error bounds for convolutional codes and asymptotically optimum decoding algorithm," *IEEE Trans. Inform. Theory*, vol. IT-13, Apr. 1967, pp. 260–269.
[13] ——, "Convolutional codes: The state diagram approach to optimal decoding and performance analysis for memoryless channels," Jet Propul. Lab., Calif. Inst. Technol., Pasadena, Space Program Summary 37-58, vol. 3, Aug. 1969, pp. 50–55.
[14] G. D. Forney, Jr., "Final report on a coding system design for advanced solar mission," Codex Corp., Watertown, Mass., NASA Contract NAS2-3637, Dec. 1967.
[15] J. K. Omura, "On the Viterbi decoding algorith," *IEEE Trans. Inform. Theory* (Corresp.), vol. IT-15, Jan. 1969, pp. 177–179.
[16] A. Gill, *Introduction to the Theory of Finite-State Machine.* New York: McGraw-Hill, 1962, pp. 7–13.
[17] G. D. Forney, Jr., "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inform. Theory*, to be published.

Single- and Multiple-Burst-Correcting Properties of a Class of Cyclic Product Codes

LALIT R. BAHL, MEMBER, IEEE, AND ROBERT T. CHIEN, MEMBER, IEEE

***Abstract*—The direct product of p single parity-check codes of block lengths $n_1,n_2,\cdots,n_p$ is a cyclic code of block length $n_1 \times n_2 \times \cdots \times n_p$ with $(n_1 - 1) \times (n_2 - 1) \times \cdots \times (n_p - 1)$ information symbols per block, if the integers $n_1,n_2,\cdots,n_p$ are relatively prime in pairs. A lower bound for the single-burst-correction (SBC) capability of these codes is obtained. Then, a detailed analysis is made for $p = 3$, and it is shown that the codes can correct one long burst or two short bursts of errors. A lower bound for the double-burst-correction (DBC) capability is derived, and a simple decoding algorithm is obtained. The generalization to correcting an arbitrary number of bursts is discussed.**

I. Introduction

A SINGLE parity-check code has one check digit per block. In the binary case, this check digit is the mod 2 sum of all the information digits. The code can also be considered as a cyclic code with generator $g(x) = (x + 1)$. The codes investigated in this paper are cyclic product codes [1] whose constituent subcodes are single parity check codes. If $n_1 < n_2 < \cdots n_p$ are integers relatively prime in pairs, then the product of p single parity check codes of block lengths $n_1,n_2,\cdots,n_p$ is a cyclic code of block length $n = n_1 \times n_2 \times \cdots \times n_p$ with $k = (n_1 - 1) \times (n_2 - 1) \times \cdots \times (n_p - 1)$ information symbols per block and having generator polynomial

$$g_p(x) = \operatorname{lcm}(x^{m_1} + 1, x^{m_2} + 1, \cdots, x^{m_p} + 1), \qquad (1)$$

where $m_u = n/n_u^{1-3}$.

The geometric structure and random error-correcting properties of these codes have been studied by Kautz [4] and Calabi and Haefeli [5], who showed that the codes have minimum distance 2^p. Burton and Weldon [1] demonstrated the cyclic nature of the codes. For $p = 2$, the codes are known as Gilbert [6] codes and their single-burst error-correction (SBC) capability has been studied by Neumann [7] and Bahl and Chien [8]. For $p > 2$, Bahl and Chien [3] have also looked at the performance of these codes with threshold decoding.

In this paper, a lower bound for the SBC capability is derived. The result is a generalization of the bound obtained for Gilbert codes by Bahl and Chien [8].

For $p = 3$, the double-burst-correction (DBC) capability is investigated, and it is shown that the codes can be used to correct one long burst or two short bursts of errors. The generalization to correcting an arbitrary number of bursts is discussed.

II. Preliminaries

In cyclic codes, the first and last digit positions of a codeword are considered to be adjacent. Therefore, there are two ways to calculate the position difference between position i_1 and position i_2 in a codeword. The two position differences are $|i_1 - i_2|$ and $n - |i_1 - i_2|$, where n is the block length. The first difference is obtained by directly traversing from i_1 to i_2, the second by an end around traverse. The smaller of the two quantities, the shortest position difference, is given by $|\bar{c}|$ where $\bar{c} \equiv (i_1 - i_2) \bmod n$ and $\bar{c}$ is taken from the set of absolutely least-magnitude residues, i.e., $\{0, \pm 1, \pm 2, \cdots, \pm(n - 1)/2\}$ if n is odd and $\{0, \pm 1, \pm 2, \cdots, \pm(n - 2)/2, n/2\}$ if n is even. This residue set is different from the more frequently used set of least nonnegative residues $\{0, 1, 2, \cdots, n - 1\}$. In this paper, it is convenient to use both residue class representations and to differentiate between the two types of residues, an overbar will be placed over least-magnitude residues.

A burst of length b is a sequence of b consecutive digits,

Manuscript received February 7, 1969; revised November 19, 1970 and February 7, 1971. This work was supported in part by National Science Foundation Grants GK 690 and GK 2339 and in part by the Joint Services Electronics Program (U.S. Army, U.S. Navy, and U.S. Air Force) under Contract DAAB-07-67-C-0199.

L. R. Bahl was with the Coordinated Science Laboratory, University of Illinois, Urbana, Ill. He is now at the IBM Thomas J. Watson Research Center, Yorktown Heights, N.Y.
R. T. Chien is with the Coordinated Science Laboratory, University of Illinois, Urbana, Ill.