A Class of Algorithms for Decoding Block Codes With Channel Measurement Information

DAVID CHASE, MEMBER, IEEE

Abstract—**A class of decoding algorithms that utilizes channel measurement information, in addition to the conventional use of the algebraic properties of the code, is presented. The maximum number of errors that can, with high probability, be corrected is equal to one less than d, the minimum Hamming distance of the code. This two-fold increase over the error-correcting capability of a conventional binary decoder is achieved by using channel measurement (soft-decision) information to provide a measure of the relative reliability of each of the received binary digits. An upper bound on these decoding algorithms is derived, which is proportional to the probability of an error for dth order diversity, an expression that has been evaluated for a wide range of communication channels and modulation techniques. With the aid of a lower bound on these algorithms, which is also a lower bound on a correlation (maximum-likelihood) decoder, we show for both the Gaussian and Rayleigh fading channels, that as the signal-to-noise ratio (SNR) increases, the asymptotic behavior of these decoding algorithms cannot be improved. Computer simulations indicate that even for low SNR the performance of a correlation decoder can be approached by relatively simple decoding procedures. In addition, we study the effect on the performance of these decoding algorithms when a threshold is used to simplify the decoding process.**

I. Introduction

BLOCK codes have been the subject of a considerable amount of research in information theory. General coding theorems [1], [2] and sophisticated implementation techniques [3]–[5] have been developed for these codes. In general, though, when dealing with binary codes, the decoding techniques developed assume a channel whose output is also binary. For many communication applications this assumption is not necessary, and furthermore, as indicated by the coding theorems developed for binary input channels with $J(> 2)$ outputs,[1] there is a significant loss in performance when just a binary output is assumed. The subject of this paper is to extend the binary decoding techniques previously developed to channels where more than just a binary output is available.

A block diagram of the communication system of interest is shown in Fig. 1. Each sequence of K binary information digits is assumed to be encoded into a block of N binary digits denoted by $X = X_1,X_2,\cdots,X_N$. These binary digits are fed into a data modulator, which determines the transmitted waveform $x(t)$. When a binary channel is assumed, the data demodulator produces a sequence of N binary digits, $Y = Y_1,Y_2,\cdots,Y_N$, which is based on the received waveform $y(t)$. In our case we shall assume that the data demodulator will supply the binary sequence Y and, in addition, a sequence of N positive numbers denoted by $\alpha = \alpha_1,\alpha_2,\cdots,\alpha_N$ will be supplied. These positive numbers, called the channel measurement information, will be used to provide a measure of the relative reliability of the received binary digits. Thus, if $\alpha_i > \alpha_j$, the decoder shall assume that Y_i is more likely to be correct than Y_j. Each value of α_i can thus be viewed as a confidence value on the reliability of each received digit. Since the decoder is fed both the sequence Y and the sequence α, we no longer have a true binary decoder.

For many applications the abstraction of channel measurement information is a relatively simple matter. For example, if we assume that the magnitude of the decision statistic of each received digit is monotonically related to the probability that the digit is received correctly, the required channel measurement information can be obtained by simply replacing a 1-bit output device (commonly a hard limiter) by a J-bit analog-to-digital converter. Thus the first (sign) bit of the analog-to-digital device yields Y_i and the remaining $J - 1$ bits specify each α_i. This type of information is also commonly referred to as soft-decision information since a decision output of more than one bit is retained.

An early example of the use of channel measurement information with block codes is given by Wagner decoding [7], [8] and its generalizations [9], where channel measurement information is used to extend by 1 the error-correcting capabilities of a code whose minimum distance is an even number. Recently considerably more sophisticated approaches for using channel measurement information with block codes have been developed [10]–[14]. Actually channel measurement information has been used to increase the efficiency of communication systems long before error-correcting codes were considered. For fading channels a common method of increasing the reliability is to transmit the same signal N times [i.e., a trivial $(N,1)$ code is employed] and combine the N analog outputs received. As has been pointed out previously [15], [16], this straight diversity combining approach is far inferior to a coding approach that utilizes channel measurements in the decoding process.

The decoding algorithms that are to be presented are attractive since they are applicable to all block codes for which a binary decoding scheme exists and can double the error-correcting capabilities of a given code. That is, for high signal-to-noise ratios (SNRs), the probability of error for a given code of minimum distance d behaves as a code

Manuscript received January 15, 1971; revised April 30, 1971. Some of the results in this paper were presented at the IEEE International Symposium on Information Theory, Noordwijk, the Netherlands, June 1970.

The author is with the General Atronics Corporation, a subsidiary of the Magnavox Company, Philadelphia, Pa. 19118.

[1] For example, if we consider the additive noise Gaussian channel for low SNRs, the improvement for $J > 2$ can be seen from the random-coding exponent [6]. For high SNRs it is necessary to use the expurgated exponent to verify the improvement possible.

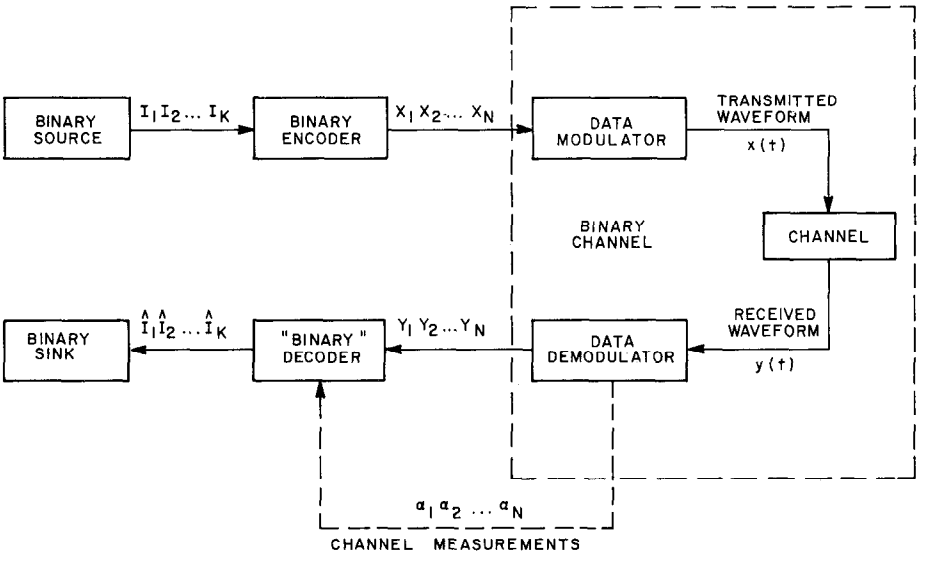

Fig. 1. Communication system block diagram.

capable of correcting $d - 1$ errors, rather than $\lfloor (d-1)/2 \rfloor$ as is possible for straight binary decoding.[2] This result, which is asymptotically the best one can do, will be shown to be true for fading channels as well as the additive noise Gaussian channel.

II. A Class of Decoding Algorithms

In this section a class of decoding algorithms that utilizes the information provided by the sequence $\boldsymbol{\alpha} = \alpha_1, \alpha_2, \cdots, \alpha_N$ is presented. These algorithms are designed to work with any binary decoder that can correct up to $\lfloor (d-1)/2 \rfloor$ errors.[3]

The binary decoder will determine the codeword $\boldsymbol{X}_m = X_{m1}, X_{m2}, \cdots, X_{mN}$, which differs in the least number of places from the received sequence $\boldsymbol{Y} = Y_1, Y_2, \cdots, Y_N$ provided that this difference is not greater than $\lfloor (d-1)/2 \rfloor$. The sequence that contains a 1 in the places where $\boldsymbol{Y}$ and $\boldsymbol{X}_m$ differ, i.e., the error sequence, is given by

$$\boldsymbol{Z}_m = \boldsymbol{Y} \oplus \boldsymbol{X}_m = Y_1 \oplus X_{m1}, Y_2 \oplus X_{m1}, \cdots, Y_N \oplus X_{mN}, \tag{1}$$

where the notation $\oplus$ represents modulo-2 addition. If we define the binary weight of a sequence $\boldsymbol{Z}_m$ as

$$W(\boldsymbol{Z}_m) = \sum_{i=1}^{N} Z_{mi}, \tag{2}$$

the function of a binary decoder is to find the codeword, or equivalently the error sequences, that satisfy

$$W(\boldsymbol{Z}_m) \leq \lfloor (d-1)/2 \rfloor. \tag{3}$$

The binary decoder will find a unique codeword if the inequality given by (3) is satisfied; otherwise, no codeword will be found. (The assumption that the binary decoder is a bounded distance decoder is for convenience only. The performance of these decoding algorithms will surely not be degraded if the binary decoder is capable of finding a codeword when (3) is not satisfied.)

A complete[4] binary decoder can be defined as a decoder capable of finding the codeword that satisfies

$$\min_m W(\boldsymbol{Y} \oplus \boldsymbol{X}_m), \tag{4}$$

where the range of m is over all possible codewords. This decoder differs from a conventional binary decoder given by (3) in that a codeword will always be obtained even if the error sequence of minimum binary weight has more than $\lfloor (d-1)/2 \rfloor$ 1s in it.

In a similar manner we can define a complete channel measurement decoder as one that is capable of finding the codeword that satisfies

$$\min_m W_\alpha(\boldsymbol{Y} \oplus \boldsymbol{X}_m). \tag{5}$$

In this case we are concerned with finding the error pattern $\boldsymbol{Z}_m = \boldsymbol{Y} \oplus \boldsymbol{X}_m$ of minimum analog weight,[5] where the analog weight of a sequence $\boldsymbol{Z}_m$ is defined as

$$W_\alpha(\boldsymbol{Z}_m) = \sum_{i=1}^{N} \alpha_i Z_{mi}. \tag{6}$$

Just as a binary decoder attempts to achieve a performance close to a complete binary decoder, the channel measurement decoders to be described will attempt to achieve a performance that is close to that for a decoder that satisfies (5). In Section III we will show that for certain

[2] $\lfloor (d-1)/2 \rfloor$ denotes the greater integer less than or equal to $(d-1)/2$.

[3] For some codes a BCH decoder [5] can correct up to $\lfloor (\tilde{d}-1)/2 \rfloor$ errors, where $\tilde{d}$ is less than the code's true minimum distance d. If such a binary decoder is used, the channel measurement decoding algorithms still apply as long as we replace d by $\tilde{d}$.

[4] This decoder is a maximum-likelihood decoder [2] when the channel is a binary symmetric channel with crossover probability $p < \frac{1}{2}$.

[5] The term "analog weight" is used to signify that the binary digits in each error pattern are weighted by their corresponding confidence values. As noted earlier, this channel measurement information need not be analog values but can be quantized to several bits.

channels the decoder given by (5) can be made equivalent to a maximum-likelihood decoder.

The basic concept behind the channel measurement decoding algorithms can be illustrated with the aid of Fig. 2. A geometric sketch is shown, which illustrates the binary distance between four codewords X_A, X_B, X_C, X_D and the received sequence Y. Each codeword is surrounded by a sphere of radius $\lfloor (d-1)/2 \rfloor$. Thus, a unique codeword, or equivalently a unique error pattern, is obtained by a binary decoder if the received sequence is within one of these spheres. In our case there is a unique error pattern $Z = Y \oplus X_A$ within the sphere of radius $\lfloor (d-1)/2 \rfloor$, which surrounds Y. The objective of the channel measurement decoder is to use a binary decoder to help obtain a relatively small set of possible error patterns rather than just one error pattern and choose the error pattern of minimum analog weight as defined by (6).

The set of error patterns considered is obtained by perturbing the received sequence Y with a test pattern T, which is a binary sequence that contains 1's in the location of the digits that are to be inverted. By adding this test pattern, modulo-2, to the received sequence a new sequence

$$Y' = Y \oplus T \tag{7}$$

is obtained and by binary decoding a new error pattern Z' is obtained. The actual error pattern relative to Y is given by

$$Z_T = T \oplus Z', \tag{8}$$

which may or may not be different from the original error pattern depending on whether or not Y' falls into the sphere of a new codeword. For the class of algorithms under consideration, the received sequence will always be perturbed within the sphere of radius $d-1$, which surrounds Y.

A flow chart for decoding with channel measurement information is shown in Fig. 3. All the decoding algorithms to be discussed will conform to Fig. 3. However, the actual set of test patterns used in each decoding algorithm will become progressively smaller. Since these decoding algorithms are to be used with a binary decoder capable of finding a codeword only when (3) is satisfied, it is possible that no legitimate error pattern will be found for a given algorithm. Under these circumstances the estimate of the codeword is given by the received binary sequence despite the fact that this sequence is surely in error.

Algorithm I

For this particular algorithm a very large set of error patterns is considered. In fact, we consider the entire set of error patterns within a sphere of radius $d-1$ about the received sequence Y shown in Fig. 2. Thus, all possible error patterns of binary weight less than or equal to $d-1$ are considered. Since the selected error pattern is determined by its analog weight, not its binary weight, it is quite possible to select an error pattern with more than $\lfloor (d-1)/2 \rfloor$ members and thus extend the error-correcting capability of the code.

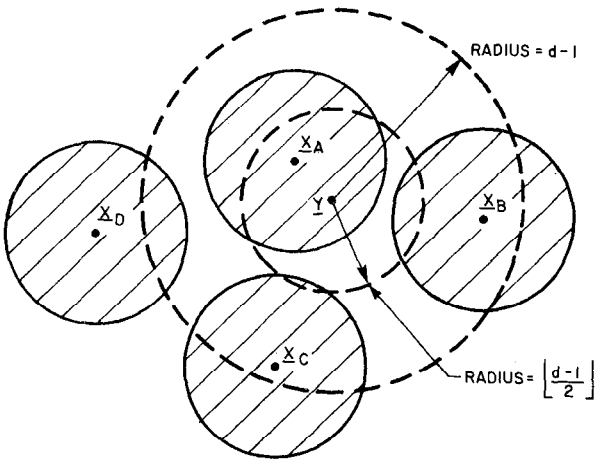

Fig. 2. Geometric sketch for decoding with channel measurement information.

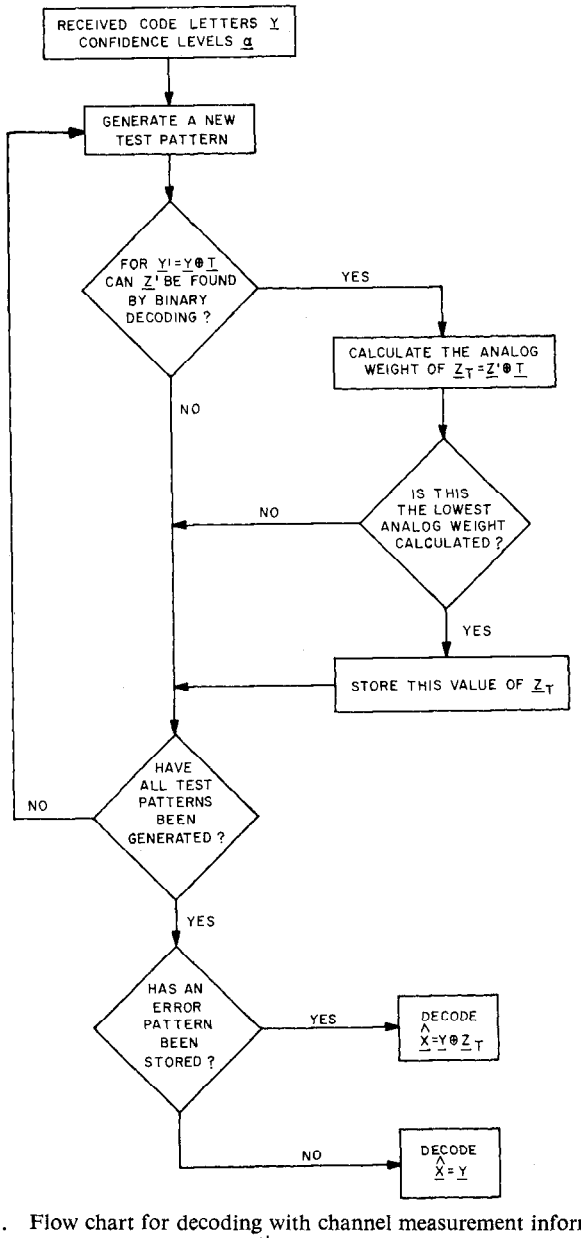

Fig. 3. Flow chart for decoding with channel measurement information.

A set of test patterns that are sufficient, but surely not necessary, to generate all error sequences of binary weight less than d are given by the set of $\boldsymbol{T}$, which contain $\lfloor d/2 \rfloor$ 1's in them. Note that the binary decoder is capable of obtaining error patterns of binary weight up to $\lfloor (d-1)/2 \rfloor$, which when combined with an appropriate test pattern can yield any error pattern that has up to $(d-1)$ members. Needless to say, since there are

$$\binom{N}{\lfloor d/2 \rfloor}$$

different test patterns, this method of implementing Algorithm 1 would only be applicable for codes whose minimum distance is quite small. Actually, a considerable reduction in the number of test patterns can be achieved by eliminating test patterns that yield identical error patterns. Nevertheless, one generally would not want to implement this particular algorithm since there are considerably simpler algorithms that perform almost as well. However, Algorithm 1 is still of interest since it can be modified (by Theorem III) to provide a lower bound on a complete channel measurement decoder and does illustrate, when compared to Algorithms 2 and 3, how the relative performance is affected by simplifications in decoding procedures.

Algorithm 2

For this algorithm a considerably smaller set of possible error patterns is considered. Only those error patterns with no more than $\lfloor (d-1)/2 \rfloor$ errors located outside the set, which contains the $\lfloor d/2 \rfloor$ lowest channel measurements, are considered. The error patterns now tested contain no more than $(d-1)$ errors, but we no longer test all possible error patterns with $(d-1)$ or less members.

A set of test patterns, which generates all required error patterns, is given by letting $\boldsymbol{T}$ have any combination of 1's, which are located in the $\lfloor d/2 \rfloor$ positions of lowest confidence values, i.e., the $\lfloor d/2 \rfloor$ positions with the lowest channel measurements. Since there are $2^{\lfloor d/2 \rfloor}$ possible test patterns, including the all-zero pattern, there are at most $2^{\lfloor d/2 \rfloor}$ error patterns considered by this decoding algorithm and generally much less. Note that if the initial error pattern $\boldsymbol{Z}$ has a binary weight of 1, then all test patterns of binary weight less than or equal to $\lfloor (d-3)/2 \rfloor$ will yield $\boldsymbol{Z}_T = \boldsymbol{Z}$.

Algorithm 3

This decoding algorithm is almost identical to Algorithm 2 except the number of test patterns considered is just $\lfloor (d/2) + 1 \rfloor$ rather than $2^{\lfloor d/2 \rfloor}$. Each test pattern has i 1's located in the i positions of lowest confidence values.[6] For a code with an even value of d, the values that i takes are given by $i = 1,3,\cdots, d-1$ and $i = 0$, which yields the all-zero test pattern $\boldsymbol{T} = \boldsymbol{0}$. When d is an odd number, $i = 0,2,4,\cdots, d-1$.

[6] This sequence of test patterns is equivalent to the sequence of erasures used in a generalized minimum distance decoding as described in [11].

This particular algorithm has the smallest set of possible error patterns and yet has the same asymptotic error performance as Algorithms 1 and 2. Computer simulated performance of these algorithms for the Golay code shows that Algorithm 3 is somewhat inferior to Algorithms 1 and 2. Nevertheless, this particular algorithm requires less computation than the previous algorithms and thus in some applications may be more desirable. This is especially true for codes whose minimum distance is large since the number of test patterns grows only linearly with the minimum distance of the code.

Decoding With a Threshold

The previous decoding algorithms can be modified to accept a given error pattern only if it satisfies a certain criterion, which is obtained by comparing the analog weight of the selected error pattern to a given threshold. The threshold that we will find useful is based on the analog weight of a test pattern $\boldsymbol{T}_k$, which has k 1's located in the positions with the k lowest confidence values. With the use of this threshold, denoted as $W_\alpha(\boldsymbol{T}_k)$, the decoding algorithms described in Fig. 3 are modified so that we decode $\hat{\boldsymbol{X}} = \boldsymbol{Y} \oplus \boldsymbol{Z}_T$ only if $W_\alpha(\boldsymbol{Z}_T) \leq W_\alpha(\boldsymbol{T}_k)$; otherwise, we decode $\hat{\boldsymbol{X}} = \boldsymbol{Y}$. The allowable range of threshold variable k is $0,1,\cdots,N$. A threshold based on $k = N$ will, of course, have no effect on a given algorithm since all error patterns must have an analog weight less than or equal to $W_\alpha(\boldsymbol{T}_N)$. On the other hand, a threshold based on $k = 0$ will have a significant effect since only error patterns of zero analog weight, i.e., a zero-weight syndrome, can be accepted.

The use of a threshold enables one to simplify a given decoding algorithm since only error patterns that satisfy the threshold criterion are accepted. For example, if Algorithm 1 is used with a threshold based on $k = \lfloor (d-1)/2 \rfloor$, only error patterns with $0,1,\cdots,\lfloor (d-1)/2 \rfloor$ (nonzero) members can satisfy the threshold. Since there is only a single error pattern with up to $\lfloor (d-1)/2 \rfloor$ members, which can be obtained by a conventional binary decoder, only the single test pattern $\boldsymbol{T} = \boldsymbol{0}$ need be used with this modified algorithm. In Section III bounds on the performance of the various decoding algorithms are obtained as a function of the value of k used in the threshold.

A further application of decoding with a threshold is obtained by noting that the data that is decoded as $\hat{\boldsymbol{X}} = \boldsymbol{Y}$ must be in error. There are various communication applications where receiving incorrect data is significantly worse then receiving no data at all. For applications of this nature as well as feedback communication links (ARQ systems) and some iterative coding schemes, the use of a threshold based on channel measurement information can be quite effective.

III. Abstraction of Channel Measurement Information

A Waveform Receiver

In Fig. 1 a model of a communication system was described. In this section we will be concerned with how

the channel measurement information is obtained from the receiver waveform given by $y(t)$.

In Fig. 4 a simplified diagram of a waveform receiver is shown. The received waveform contains information on values of the N binary digits in the coded sequence $X_1,X_2,\cdots,X_N$. This information is abstracted from $y(t)$ by the N decision filters shown on the figure. The output of these decision filters is a weighted decision statistic. The decision statistic, denoted as v_i, is assumed to be of a form such that the decision rule for a given digit X_i is

$$\begin{aligned} &\text{if } v_i \geq 0, \qquad \text{say } \hat{X}_i = Y_i = 0 \\ &\text{if } v_i < 0, \qquad \text{say } \hat{X}_i = Y_i = 1. \end{aligned} \tag{9}$$

The output of these filters is not just a decision statistic, but actually is a decision statistic multiplied by a positive weight factor given by w_i. The channel measurement or confidence value for each received binary digit Y_i is defined as the magnitude of the output of the decision filter, and thus is given by

$$\alpha_i = |w_i v_i|. \tag{10}$$

The exact form of these decision filters depends on the form of the transmitter signals and the type of channel that the signals are transmitted over. The analysis of the decoding algorithm is in terms of these weighted decision statistics and thus applies for any situation where the waveform receiver can be cast in the form shown in Fig. 4. Of course, if the decoding algorithm is to perform successfully, these decision filters must be designed in a manner such that the magnitude of the weighted decision statistics represents meaningful confidence values on the received binary digits.

Binary Antipodal Signaling Over the White Gaussian Noise Channel and the Rayleigh Fading Channel

For the Gaussian channel an estimate of X_i is obtained from the ith decision filter, which is assumed to be matched to a waveform of energy E, which is used to represent the value of X_i. The output of this matched filter is

$$w_i = 1 \qquad v_i = \begin{cases} \sqrt{E} + n_i, & X_i = 0, \\ -\sqrt{E} + n_i, & X_i = 1, \end{cases} \tag{11}$$

where n_i is a component of additive noise Gaussian process whose (single-sided) noise power per hertz is given by N_0. The variance of n_i is

$$\overline{n_i^2} = \frac{N_0}{2} \tag{12}$$

and the SNR for this channel is defined as

$$\gamma = \frac{E}{2\overline{n_i^2}} = \frac{E}{N_0}. \tag{13}$$

In this case the decision filters have been chosen such that complete channel measurement decoding as defined by (5) is equivalent to maximum-likelihood decoding [6].

To prove this assertion, we first note that for the white Gaussian channel the decision rule for a correlation decoder given by

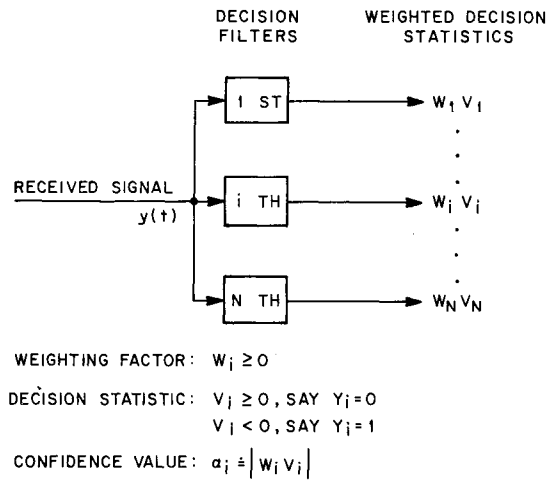

Fig. 4. Waveform receiver.

$$\max_m \sum_{i=1}^{N} w_i v_i (-1)^{X_{mi}} \tag{14}$$

is equivalent to selecting the codeword that maximizes $P(\boldsymbol{WV} \mid \boldsymbol{X}_m)$. In general, for a given channel the decision rule given by (14) may not be the maximum-likelihood estimate but nevertheless is applicable to any waveform receiver of the form shown in Fig. 4.

To relate the above decision rule for a correlation decoder to that given by (5), (14) is first divided up into a summation over those i such that $X_{mi} \neq Y_i$, denoted as the set S_m, and those i such that $X_{mi} = Y_i$. Thus, the decision rule given by (14) can be written as

$$\max_m \Big[\sum_{i \notin S_m} |w_i v_i| - \sum_{i \in S_m} |w_i v_i|\Big]. \tag{15}$$

We can now rewrite (15) in terms of α_i given by (10). Furthermore, we can sum the first term over all i by subtracting off those terms in the set S_m that are added to give

$$\max_m \Big[\sum_{\text{all } i} \alpha_i - 2 \sum_{i \in S_m} \alpha_i\Big]. \tag{16}$$

Since the first summation term above is the same for all codewords, maximizing (16) is equivalent to minimizing the following expression

$$\min_m \sum_{i \in S_m} \alpha_i = \min_m \sum_{i=1}^{N} \alpha_i Z_{mi} \doteq \min_m W_\alpha(\boldsymbol{Z}_m), \tag{17}$$

where the right-hand side of (17) follows from the observation that $\boldsymbol{Z}_m = \boldsymbol{Y} \oplus \boldsymbol{X}_m$ has a 1 in those places where $i \in S_m$ and $Z_{mi} = 0$ for $i \notin S_m$. Thus, we have shown that the decision rules based on (5) and (14) are equivalent for any channel and, furthermore, for the white Gaussian channel, complete channel measurement decoding is equivalent to maximum-likelihood decoding.

For the Rayleigh fading channel we assumed that the waveform used to represent X_i is perturbed by (Rayleigh-distributed) multiplicative noise denoted as r_i in addition to the additive white Gaussian noise. Assuming the mag-

nitude of the Rayleigh noise (as well as the phase of the received signal) is known, the output of the ith decision filter when chosen to maximize $P(\boldsymbol{WV} \mid \boldsymbol{X}_m)$ is given by

$$w_i = r_i \qquad v_i = \begin{cases} r_i\sqrt{E} + n_i, & X_i = 0 \\ -r_i\sqrt{E} + n_i, & X_i = 1. \end{cases} \quad (18)$$

The average SNR for this case is still given by (13) when the multiplicative noise is normalized such that $\overline{r_i^2} = 1$.

The general problem of choosing the decision filters and analyzing their performance has been pursued in the past [17]–[23] by those concerned with the reception of (uncoded binary) signals over a multichannel or channel where there are several received noisy versions of the transmitted signal, such as is the case for diversity reception. For multichannel reception the outputs of the decision filters are summed (diversity combined) in order to obtain a single decision statistic for the transmitted signals. For the reception of a block of coded binary data, as in our case, the outputs of these decision filters are viewed as the inputs to a channel measurement decoder. We will be able to show that an upper bound on the decoding algorithm can be obtained in terms of the error probability for multichannel (diversity) reception. This result is quite useful since there are many articles, the previous references [17]–[23] being just a sample, on evaluating the error probability for multichannel reception for a wide range of communication channels and modulation–demodulation techniques.

The results of this section are summarized as follows.

Lemma 1: The decision rule for a complete channel measurement decoder given by

$$\min_m \sum_{i=1}^{N} \alpha_i Z_{mi}$$

and that of a correlation decoder given by

$$\max_m \sum_{i=1}^{N} w_i v_i (-1)^{X_{mi}}$$

are equivalent and applicable to any waveform receiver of the form shown in Fig. 4. For the Gaussian and Rayleigh fading channel a waveform receiver can be chosen such that the above decision rules are equivalent to maximum-likelihood decoding.

IV. Theorems on Decoding With Channel Measurement Information

In this section several theorems on the probability of error for decoding with channel measurement information are given. In determining an upper bound on the probability of a block error for a group code, we can, without loss in generality, assume that the transmitted codeword is the all-zero sequence [3], [6]. Thus, if the weighted decision statistic $w_i v_i$ is negative, the received binary digit Y_i is not equal to X_{mi}. The upper bounds to be presented in this section will be written in a form that assumes that the letters X_{mi} of the transmitted codeword $\boldsymbol{X}_m$ are all zero.

The fundamental theorem of this section, which is proven in Appendix I-A, is as follows.

Theorem 1: The probability of a block error for a code of minimum distance d and block length N is upper-bounded by

$$\binom{N}{d}$$

times the probability that a set of d weighted decision statistics has a negative sum. Since this probability is also the probability of error for a system with dth-order diversity, this upper bound can be written as

$$P_e(\text{block}) < \binom{N}{d} \Pr\left[\sum_{i=1}^{d} w_i v_i < 0\right] = \binom{N}{d} P_e(d\text{th-order diversity}). \quad (19)$$

This upper bound is valid for Algorithms 1–3 as well as for complete channel measurement decoding, which by Lemma 1, is equivalent to correlation decoding.

The upper bound given by Theorem 1 applies to all waveform receivers of the form given in Fig. 4, providing that $P_e(d\text{th-order diversity})$ is the same for all possible sets of d weighted decision statistics. A sufficient condition for the above to be true is that the set of $w_i v_i$ be statistically independent. Assuming statistical independence the bound given by Theorem 1 can be simplified further by the following lemma.

Lemma 2: The probability of an error for dth-order diversity can be upper-bounded by

$$P_e(d\text{th-order diversity}) < \langle e^{-\lambda wv} \rangle^d, \qquad \text{for any } \lambda > 0 \quad (20)$$

for statistically independent weighted decision statistics. The angular brackets are used to denote the average over the weighted decision statistic wv.

The proof of Lemma 2 follows directly from the application of a Chernoff bound [2], [6].

In Section II we introduced the concept of modifying Algorithms 1–3, as well as correlation decoding, by accepting an error pattern only if its analog weight is less than or equal to a threshold based on the sum of the lowest confidence values, denoted as $W_\alpha(\boldsymbol{T}_k)$. In Appendix I-B the following theorem on decoding with a threshold is proven.

Theorem 2: The probability of a block error on decoding a code of minimum distance d and block length N with a threshold given by $W_\alpha(\boldsymbol{T}_k)$ is upper-bounded by

$$P_e(\text{block}) < \binom{N}{d} \Pr\left[\sum_{i=1}^{d} w_i v_i < 0\right], \quad \text{for a threshold based on } k = d-1, d, d+1, \cdots, N \quad (21)$$

and

$$P_e(\text{block}) < \binom{N}{k+1} \Pr\left[\sum_{i=1}^{k+1} w_i v_i < 0\right], \quad \text{for a threshold based on } k = 0, 1, \cdots, d-1. \quad (22)$$

These upper bounds are valid when a threshold is used for Algorithms 1–3 and correlation decoding.

From these bounds we note that a threshold based on $k \geq d - 1$ yields the same upper bound as that given in Theorem 1. Thus, the simplifications obtained in decoding by only considering error patterns of analog weight less than or equal to $W_\alpha(\boldsymbol{T}_{d-1})$ does not affect the upper bound previously derived. On the other extreme, for a threshold based on $k = 0$, where only an error pattern of zero analog weight is accepted, we obtain a standard upper bound on a block of N digits given by

$$P_e(\text{block}) < Np, \tag{23}$$

where

$$p = \Pr[w_i v_i < 0] \tag{24}$$

is the probability of a bit error before decoding. This is as expected since a threshold of zero does not allow for any error correction regardless of the strength of the code.

An interesting property of Algorithm 1 is obtained when this algorithm is used with a threshold of $W_\alpha(\boldsymbol{T}_{d-1})$. If an error pattern satisfying the threshold is selected by this algorithm, it is exactly the same error pattern that would be selected by a correlation decoder. Note that those error patterns not considered by Algorithm 1, but considered by correlation decoding, must have an analog weight greater than $W_\alpha(\boldsymbol{T}_{d-1})$ since they are patterns with more than $d - 1$ (nonzero) members. If we define the undetected block errors as those block errors where the selected error patterns have an analog weight less than or equal to the threshold, we can write the following theorem.

Theorem 3: The probability of an undetected block error for Algorithm 1 used with a threshold of $W_\alpha(\boldsymbol{T}_{d-1})$ yields a lower bound on the probability of error for a correlation decoder.

This theorem follows directly from the observation that the calculation of probability of an undetected error for Algorithm 1 only includes those block errors that are also made by a correlation decoder and does not include those block errors made by a correlation decoder when the analog weight of the selected error pattern is greater than $W_\alpha(\boldsymbol{T}_{d-1})$.

A lower bound on channel measurement decoding, which will be useful in determining the asymptotic behavior of these decoding algorithms, is obtained by noting that a block error will be made if the received binary sequence is exactly equal to an incorrect codeword. Under these conditions there will exist one incorrect error pattern of zero analog weight and thus a block error will be made for Algorithms 1–3 and correlation decoding. The lower bound given by the probability of this event can be written in terms of $d(m,m')$, the binary distance between $\boldsymbol{X}_m$, the transmitted codeword, and another codeword $\boldsymbol{X}_{m'}$, as

$$P_e(\text{block}) > \sum_{m' \neq m} p^{d(m,m')}(1 - p)^{N - d(m,m')}, \tag{25}$$

where, as noted previously, p is the probability of a bit error before decoding.

For a code of minimum distance d we shall use n_d to denote the number of codewords that are exactly a distance d away from the transmitted codeword. Note that for group code n_d is the number of codewords of weight d. In terms of n_d a lower bound on (25) can be obtained by including only those $m' \neq m$ with $d(m,m') = d$, which gives the following theorem.

Theorem 4: The probability of block error for a code of minimum distance d, with n_d codewords of weight d, can be lower-bound by

$$P_e(\text{block}) > n_d p^d (1 - p)^{N-d}. \tag{26}$$

This lower bound is valid for correlation decoding, Algorithms 1–3, and also is a lower bound on the probability of an undetected error when any of the above decoding procedures are used with a threshold of $W_\alpha(\boldsymbol{T}_k)$, where $k = 0,1,\cdots,N$.

The bound given by (26) is quite loose in that it is even valid for the probability of an undetected block error for a threshold of zero. In this case a block error can only be made if the selected error pattern is incorrect and has zero weight, or equivalently, the received binary sequence is exactly equal to an incorrect codeword.

The theorems derived thus far are fairly general in that they apply to any waveform receiver of the form given on Fig. 4 and apply regardless of how useful the actual channel measurement information is. In the remainder of this section we shall evaluate the bounds derived for the specific waveform receivers discussed in Section III. The effectiveness of the channel measurement decoding algorithms will be demonstrated by first comparing these bounds with the performance of a maximum-likelihood decoder and finally comparing the performance to that of a conventional binary decoder.

A. Performance for the White Gaussian Noise Channel

For the Gaussian channel the weighted decision statistic is given by (11) and the SNR per transmitted letter is given by (13). In order to account for the redundancy in codes of different rates, we shall use the SNR per transmitted bit of information in our bounds. This normalized SNR is given in terms of the data rate $R = K/N$ as

$$\gamma_b \doteq \frac{E_b}{N_0} = \gamma \frac{N}{K} = \frac{\gamma}{R}. \tag{27}$$

In terms of γ_b we can write the bit error given by (24) as

$$p = \int_{\sqrt{2\gamma_b R}}^{\infty} \left(\frac{e^{-x^2/2}}{\sqrt{2\pi}}\right) dx \doteq Q(\sqrt{2\gamma_b R}). \tag{28}$$

With the aid of (28), the lower bound given by Theorem 4 can immediately be written in terms of γ_b as

$$P_e(\text{block}) > n_d Q(\sqrt{2\gamma_b R})^d (1 - Q(\sqrt{2\gamma_b R}))^{N-d}. \tag{29}$$

For the Gaussian channel as the SNR increases the asymptotic behavior of this lower bound can readily be shown to be of the form

$$P_e(\text{block}) \sim \exp\{-\gamma_b[Rd + o(\gamma_b)]\}, \tag{30}$$

where $o(\gamma_b)$ is defined as a function that goes to zero as $\gamma_b \to \infty$.

To evaluate the upper bound given by (19) in Theorem 1, we must evaluate

$$P_e(d\text{th-order diversity}) = \Pr\left[\left(d\sqrt{E} + \sum_{i=1}^{d} n_i\right) < 0\right], \tag{31}$$

where for the white noise channel each Gaussian variable is identically distributed and statistically independent. The above probability is given by

$$P_e(d\text{th-order diversity}) = Q(\sqrt{2d\gamma_b R}) \tag{32}$$

with an asymptotic behavior given by (30).

These results yield the following theorem.

Theorem 5: For the white Gaussian noise channel the probability of an error for Algorithms 1–3 and correlation decoding is bounded by

$$n_d Q(\sqrt{2\gamma_b R})^d (1 - Q(\sqrt{2\gamma_b R}))^{N-d} < P_e(\text{block}) < \binom{N}{d} Q(\sqrt{2d\gamma_b R}). \tag{33}$$

As the SNR increases, the asymptotic behavior of the upper bound and lower bound is the same and is given by

$$P_e(\text{block}) \sim K \exp(-\gamma_b d R), \tag{34}$$

where $K = \exp[-\gamma_b o(\gamma_b)]$, which is a "weak function" of γ_b.

Since the lower bound given by (26) is also a lower bound on a correlation or, in this case, a maximum-likelihood decoder, we have the following corollary.

Corollary 1: For the white Gaussian channel the asymptotic behavior of the probability of error for Algorithms 1–3 is the same as that for a maximum-likelihood decoder.

Another corollary, which follows directly from Theorem 2, is as follows.

Corollary 2: For the white Gaussian noise channel the results of Theorem 5 and Corollary 1 also apply when a threshold $W_\alpha(\boldsymbol{T}_k)$, with $k \geq d - 1$, is used for Algorithms 1–3 and correlation decoding.

B. Performance for the Coherent Rayleigh Fading Channel

For the fading channel the weighted decision statistics are given by (18) and the normalized SNR is given by (27).

In order to evaluate the lower bound given in Theorem 4, we need to calculate the value of p in (24), which is [18], [23]

$$p = \tfrac{1}{2}[1 - \mu], \tag{35}$$

where the value of μ in the above expression is given by

$$\mu = \left(\frac{\gamma_b R}{\gamma_b R + 1}\right)^{1/2}. \tag{36}$$

Thus, the lower bound on $P_e(\text{block})$ is

$$P_e(\text{block}) > n_d \left(\frac{1-\mu}{2}\right)^d \left(\frac{1+\mu}{2}\right)^{N-d}, \tag{37}$$

which for high SNR behaves as

$$P_e(\text{block}) \sim n_d \left(\frac{1}{4\gamma_b R}\right)^d. \tag{38}$$

The above expression decreases algebraically with the SNR rather than exponentially, as was the case for the nonfading Gaussian channel. The expression given in (38) can be rewritten as

$$P_e(\text{block}) \sim \exp\{-\ln \gamma_b [d + o(\gamma_b)]\}, \tag{39}$$

where $o(\gamma_b)$ is a function that goes to zero as $\gamma_b \to \infty$.

For the upper bound of Theorem 1 we must evaluate the probability of an error for dth-order diversity. This expression is given by [18], [23]

$$P_e(d\text{th-order diversity}) = \frac{1}{2}\left[1 - \mu \sum_{k=0}^{d-1} \binom{2k}{k} \left(\frac{1-\mu^2}{4}\right)^k\right] \tag{40}$$

for the case where the Rayleigh variables are statistically independent. The value of μ above is given by (36).[7] The asymptotic form of the above expression can be difficult to obtain [24] if Lemma 2 is not used. With the aid of Lemma 2 we can write

$$P_e(d\text{th-order diversity}) < \left(\frac{1}{1 + \lambda E^{1/2} - (\lambda^2 N_0)/4}\right)^d, \tag{41}$$

where the tightest bound is obtained by choosing $\lambda = 2E^{1/2}/N_0$ to give

$$P_e(d\text{th-order diversity}) < \left(\frac{1}{1+\gamma_b R}\right)^d < \left(\frac{1}{\gamma_b R}\right)^d. \tag{42}$$

Thus the probability of a block error can be upper-bounded by

$$P_e(\text{block}) < \binom{N}{d} \left(\frac{1}{\gamma_b R}\right)^d, \tag{43}$$

which can be written in the form given by (39).

These results yield the following theorem.

Theorem 6: For the Rayleigh fading channel the probability of an error for Algorithms 1–3 and correlation decoding is bounded by

$$n_d \left(\frac{1-\mu}{2}\right)^d \left(\frac{1+\mu}{2}\right)^{N-d} < P_e(\text{block}) < \frac{1}{2}\binom{N}{d}\left[1 - \mu \sum_{k=0}^{d-1} \binom{2k}{k}\left(\frac{1-\mu^2}{4}\right)^k\right], \tag{44}$$

where μ is given by (36). As the SNR increases, the asymptotic behavior of the upper and lower bounds is the same to within a multiplicative constant and is given by

$$P_e(\text{block}) \sim K \exp(-\ln \gamma_b d) = K(1/\gamma_b)^d. \tag{45}$$

The coefficient $K = \exp[-o(\gamma_b) \ln(\gamma_b)]$ is in this case not a function of the γ_b.

[7] For several other forms of modulation the expression given by (40) is valid as long as the value of μ is redefined. As an example for binary differential phase-shift keying (DPSK) the value of μ is [23] $(\gamma_b R)/(\gamma_b R + 1)$.

Corollary 3: For the Rayleigh fading channel the asymptotic behavior of the probability of error for Algorithms 1–3 exhibits the same behavior with increasing SNR as a maximum-likelihood decoder.

Also from Theorem 2 we have the following corollary.

Corollary 4: For the Rayleigh fading channel the results of Theorem 6 and Corollary 3 apply for Algorithms 1–3, and correlation decoding when a threshold given by $W_\alpha(\boldsymbol{T}_k)$, $k \geq d - 1$, is used.

C. Performance Comparison With Conventional Binary Decoding

For conventional binary decoding the probability of a block error for a code of minimum distance d is given by

$$P_e(\text{block}) = \sum_{i=e+1}^{N} \binom{N}{i} p^i (1-p)^{N-i}, \tag{46}$$

where

$$e = \begin{cases} (d/2) - 1, & d \text{ even} \\ (d-1)/2, & d \text{ odd.} \end{cases} \tag{47}$$

The probability of a raw bit error is given by p and

$$\binom{N}{i}$$

is the number of error patterns with i errors that cannot be corrected. For conventional binary decoding all error patterns with e errors or less are corrected and no error patterns with more than e errors are corrected.

For high SNR the first term in the summation dominates and thus for the Gaussian channel we have

$$P_e(\text{block}) \sim K \exp\{-\gamma_b[R(e+1)]\}. \tag{48}$$

Similarly for the coherent Rayleigh fading channel we have

$$P_e(\text{block}) \sim K(1/\gamma_b)_{(e+1)}. \tag{49}$$

Comparing the asymptotic forms above with those given by (34) and (45), we see that for high SNR decoding with channel measurement information effectively doubles the error-correcting capabilities of the particular code under consideration. More precisely, in the asymptotic form of $P_e(\text{block})$ for Algorithms 1–3, the term $e + 1$ is replaced by the code's minimum distance d. Thus, the use of channel measurement information extends the effective error-correcting capability of the code from e given by (47) to

$$e_{\text{channel measurement decoding}} = d - 1.$$

This increase in the code's error-correcting capabilities results in a change in the exponent by the factor $d/(e+1)$. Thus, for d even an asymptotic improvement in the SNR for the Gaussian channel of 3 dB is achieved by decoding with channel measurement information.[8] For the Rayleigh fading channel the use of channel measurement information can result in a SNR improvement considerably larger than 3 dB, since the error probability varies algebraically with γ_b and changes from $(e+1)$th-order diversity behavior to the behavior of a dth-order diversity system.

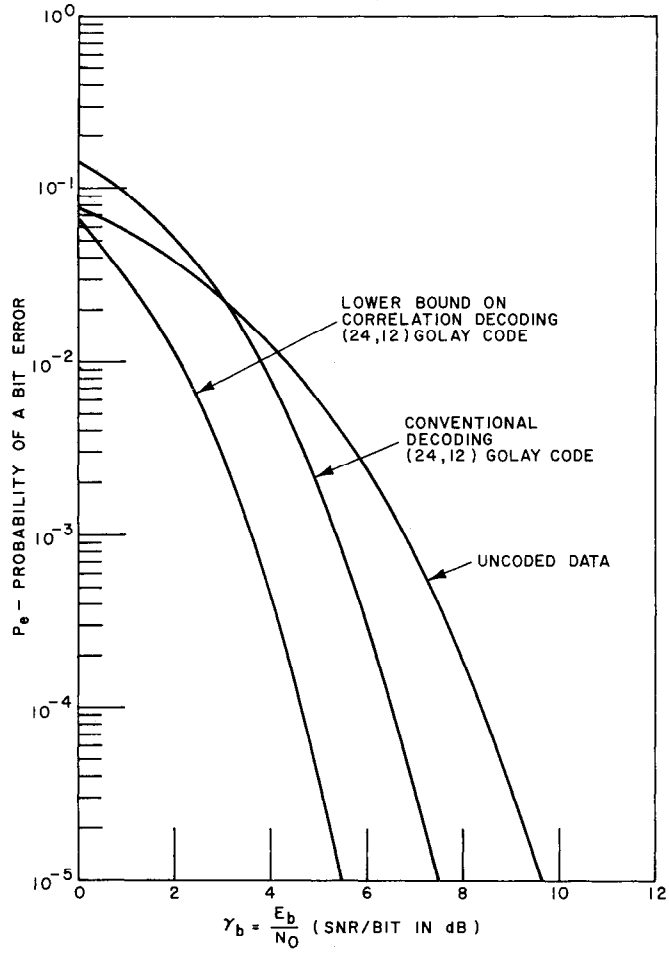

Fig. 5. Performance curves for the Gaussian channel (2ϕ – *PSK*).

V. Computer Simulated Results

In the previous section the bounds derived were useful in determining the asymptotic behavior of the error probability, but were not tight enough to distinguish the performance differences for Algorithms 1–3. In this section we compare by computer simulations the performance of the various decoding algorithms for binary antipodal signals over the Gaussian and Rayleigh fading channels. The code chosen is the distance 8 (24,12) extended Golay code.

To make the comparisons more meaningful, we first obtain three reference curves for both channels. These curves are the probability of a bit error for uncoded data, the probability of a bit error for conventional binary decoding of an extended Golay code, i.e., all patterns of one, two, and three errors are corrected and patterns of four errors are detected but not corrected, and a tight lower bound on the probability of bit error for a correlation decoder. This lower bound is actually the probability of an undetected bit error for Algorithm 1 used with a threshold

[8] It is interesting to note that for high SNRs this 3-dB performance gain does not show up in the random coding exponent [6], but can be verified from the expurgated exponent [2], which is a considerably tighter bound for $\gamma_b \to \infty$.

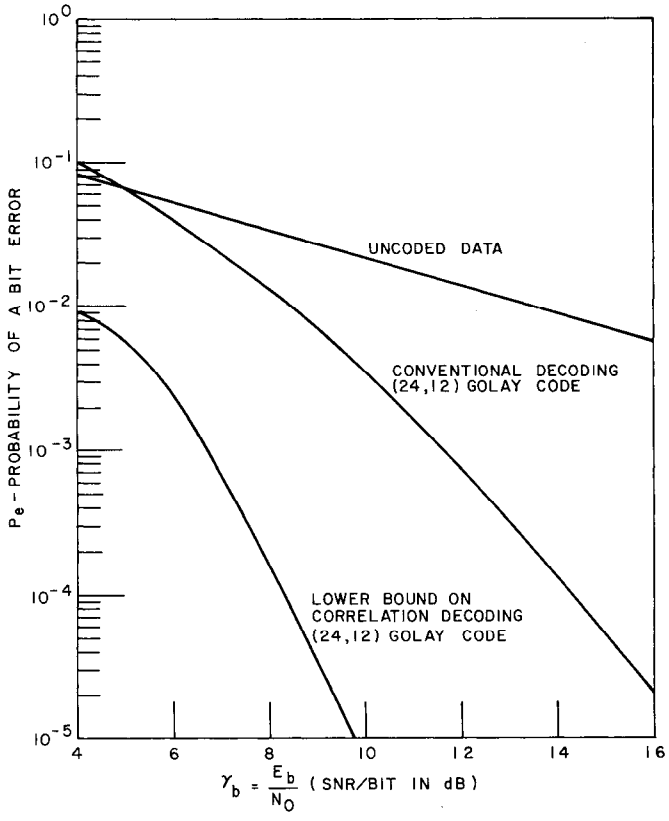

Fig. 6. Performance curves for the Rayleigh fading channel (2ϕ – PSK).

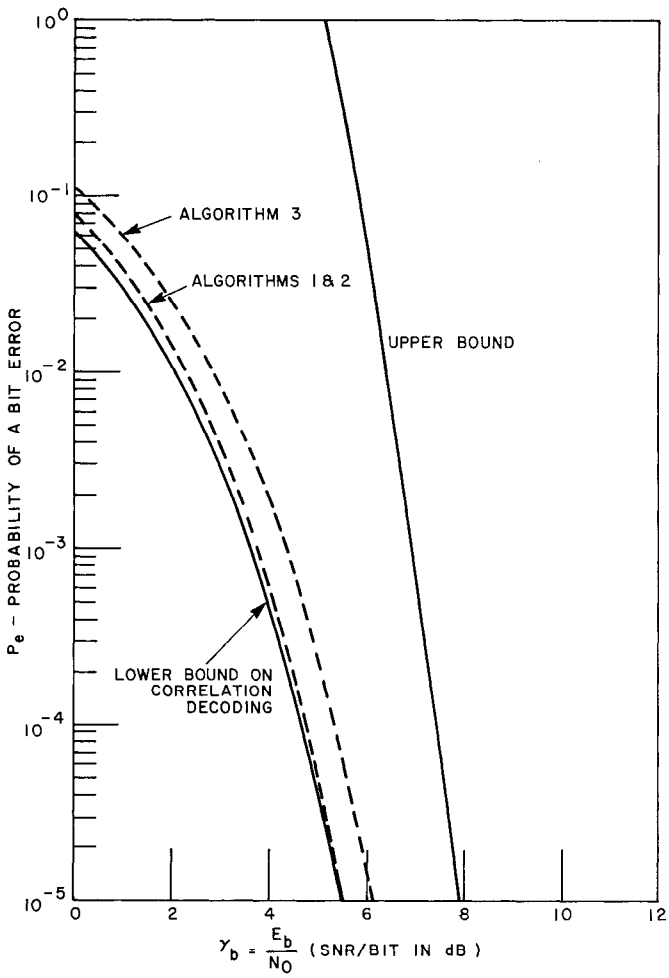

Fig. 7. Performance of a (24,12) Golay code over the Gaussian channel (2ϕ – PSK).

of $W_\alpha(T_7)$, which by Theorem 3 is a valid lower bound. The behavior of these reference curves as a function of γ_b are shown on Figs. 5 and 6 for the Gaussian channel and the Rayleigh fading channel, respectively. It is noted that the probability of a bit error on a decoded block of data is plotted, rather than P_e(block), so that a fair comparison can be made between an uncoded and coded signaling scheme.

From Fig. 5 we see that at $P_e = 10^{-5}$, conventional binary decoding offers a 2-dB saving in power and correlation decoding offers a 4-dB improvement. Asymptotically the binary decoding gain is 3 dB [note that the exponent in (48) is $R(e + 1)\gamma_b = 2\gamma_b$], and the channel measurement decoding gain is 6 dB [exponent in (34) is $Rd\gamma_b = 4\gamma_b$]. For the fading channel shown in Fig. 6 the corresponding savings in power are considerably larger, as can be seen by comparing the slopes of the three reference curves. For high SNR the probability of error decreases at a rate of one order of magnitude per 10 dB for uncoded data, per 2.5 dB for binary Golay decoding, and per 1.25 dB for correlation decoding.

On Fig. 7 the three decoding algorithms are compared for the Gaussian channel. The lower bound on correlation decoding from Fig. 5 is included and the theoretical upper bound on P_e(block) given by (33), which is also an upper bound on P_e, is included. From this figure we note that the performance of Algorithms 1 and 2 are indistinguishable even though Algorithm 1 is considerably more complex. Note that for Algorithm 2 only $2^{\lfloor d/2 \rfloor} = 16$ test patterns are used. Furthermore, we have the somewhat surprising result that performance of a correlation decoder is achieved by Algorithms 1 and 2 for SNR almost as low as 0 dB. The performance of Algorithm 3, while asymptotically optimum, is somewhat inferior to the previous algorithms. The corresponding curves for the fading channel shown in Fig. 8 illustrate a larger spread between the performance of the various decoding algorithms. The theoretical upper bound on this figure given by (44) does not provide a good estimate of the error probability, but its slope indicates that the error probability for decoding a Golay code with channel measurements decreases at a rate of one order of magnitude per 1.25 dB, as in the case for eighth-order diversity.

Acknowledgment

The author would like to thank J. S. Zaborowski, Jr., for the computer simulations that were presented.

Appendix I

A. *Proof of Theorem 1*

The upper bound on P_e(block) given by Theorem 1 is valid for Algorithms 1–3 and correlation decoding. The proof for each of the above four cases is somewhat different and thus each case is treated

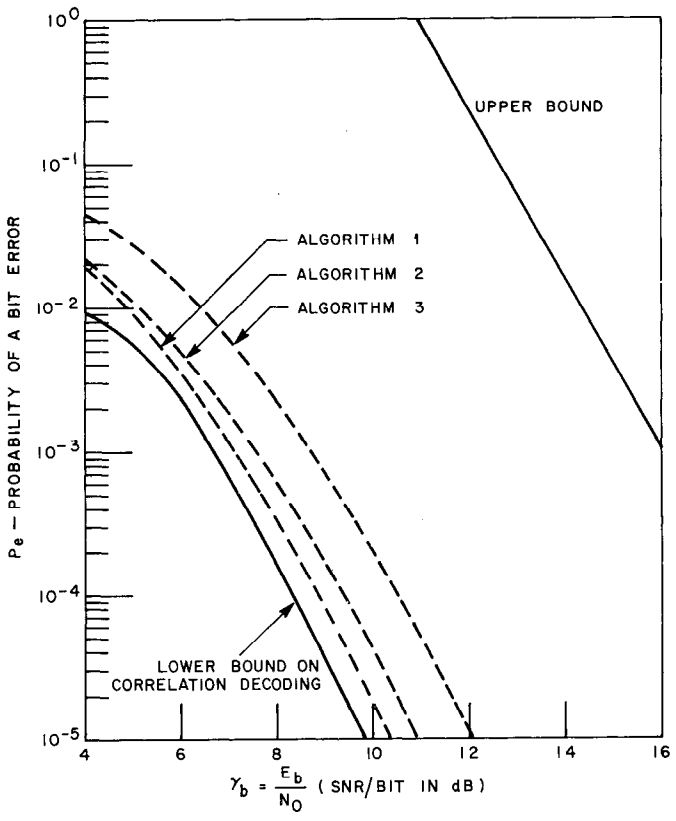

Fig. 8. Performance of a (24,12) Golay code over the Rayleigh fading channel (2ϕ — PSK).

separately. In general, we will show for all four possible cases that if a block error occurs, there exists a set S_d of d weighted decision statistics such that

$$\sum_{i \in S_d} w_i v_i < 0, \tag{50}$$

where the transmitted codeword is assumed to be the all-zero sequence, i.e., a negative value of $w_i v_i$ indicates an error before decoding. Once we show that condition (50) is a necessary condition for a block error, we can write

$$P_e(\text{block}) < \Pr\left[\sum_{i \in S_d} w_i v_i < 0 \text{ for some set } S_d\right]. \tag{51}$$

Since there are

$$\binom{N}{d}$$

different sets of d weighted decision statistics, (51) can be upper-bounded further by bounding the probability of a union of events by their sum to give

$$P_e(\text{block}) < \binom{N}{d} \Pr\left[\sum_{i=1}^{d} w_i v_i < 0\right]. \tag{52}$$

In (52) we assume that all sets of d decision statistics have the same probability of having a negative sum.

To prove Theorem 1, we begin with the shortest proof that applies to correlation decoding.

1) Correlation Decoding: For this case an error will be made if the analog weight of the true error pattern denoted as $W_\alpha(Z)$ is greater than the analog weight of some other error pattern denoted as $W_\alpha(Z')$. If we denote $S(Z)$ as the set i for which the letters of Z are equal to 1, i.e., the locations of the true error pattern, and $S(Z')$ as the set i giving the location of the errors in the pattern Z', we can write the above condition as

$$\sum_{i \in S(Z')} |w_i v_i| < \sum_{i \in S(Z)} |w_i v_i|. \tag{53}$$

For all $i \subset S(Z)$ the weighted decision statistics are negative so that from (53) we can obtain the following negative sum:

$$\sum_{i \in S(Z')} |w_i v_i| + \sum_{i \in S(Z)} w_i v_i < 0. \tag{54}$$

Since the code has a minimum distance of d, the number of different decision statistics in the union of sets $S(Z')$ and $S(Z)$ must be at least d. Note that the number of 1's in $Z' \oplus Z$ is the binary distance between two distinct codewords. Now if we retain those $w_i v_i$ contained in $S(Z')$ and not contained in $S(Z)$ [i.e., those $i \in S(Z') \cap \bar{S}(Z)$], the left side of (54) can be lower-bounded by the following negative summation:

$$\sum_{i \in S(Z') \cap S(Z)} w_i v_i + \sum_{i \in \bar{S}(Z)} w_i v_i < 0, \tag{55}$$

which contains at least d terms. Since the d lowest terms in (55) must form a negative sum, we have shown that (50) is a necessary condition for an error when correlation decoding is used.

2) Algorithm 1: For this algorithm only those error patterns of binary weight less than d are tested. Now a block error will be made if one of the following two events occur.

i) The true error pattern is considered and its analog weight is higher than some other error pattern considered by the given algorithm.

ii) The true error pattern is not considered by the given algorithm.

If a block error is made because of the first condition, i.e., (53) is true, the proof of Theorem 1 is the same as given in 1).

In the second case a block error is made because the true error pattern is not considered. However, since all error patterns with less than d members are tested, the true error pattern must have at least d members. By retaining d members of the true error pattern that contains all negative values of $w_i v_i$, we can satisfy condition (50) and hence Theorem 1.

3) Algorithm 2: As was true for Algorithm 1, there are two events for a block error.

The proof for the case when the true error pattern is considered and its analog weight is not the lowest is the same as given previously.

For the case when the true error pattern is not tested, there must exist at least $\lfloor (d+1)/2 \rfloor$ errors, i.e., negative values of $w_i v_i$, outside the set containing the $\lfloor d/2 \rfloor$ lowest channel measurements. The combined set of weighted decision statistics made up of the $\lfloor d/2 \rfloor$ lowest values of $|w_i v_i|$ and $\lfloor (d+1)/2 \rfloor$ errors outside of this set must form a negative sum satisfying (50). Note that the pairwise sum of an error outside the set of $\lfloor d/2 \rfloor$ lowest channel measurements and a member within this set must form a negative sum. For d even there exist $d/2$ of these pairwise negative sums, which, when combined, yields a net negative sum containing d members. For d odd we use $(d-1)/2$ of the available $(d+1)/2$ errors to form $(d-1)/2$ pairwise negative sums and a single negative value of $w_i v_i$ to satisfy (50).

4) Algorithm 3: Of the two possible events for a block error the case where the true error pattern is considered, but not selected, is treated as previously. To show that (50) is also satisfied for the remaining event, i.e., when the true error pattern is not tested, is somewhat more involved. For simplicity we will restrict the proof of this event to the case where the code's minimum distance is even. The case where d is odd can be treated in a very similar manner.

Our goal is to show that (50) holds when the true error pattern is not obtained by perturbing the received binary sequence Y with test patterns $T_0, T_1, T_3, \cdots, T_{d-1}$, where the notation T_i is used to represent a test pattern that has i 1's in the i positions of lowest confidence values.

We start by assuming that the true error contains exactly $d/2$ errors and is not tested by Algorithm 3. Note that if the true error pattern has less than $d/2$ errors, it must be tested.

Since the binary decoder is capable of finding error patterns with

up to $\lfloor (d-1)/2 \rfloor$ errors, the true error pattern with $d/2$ errors will not be obtained by $Y \oplus T_0$.

If the error pattern is not obtained when Y is perturbed by T_1, there must exist $d/2$ errors outside the set containing the lowest confidence value denoted as $S(T_1)$. Note that if a member of the true error pattern was contained in the set $S(T_1)$, then only $(d/2) - 1$ errors would be contained in the sequence $Y \oplus T_1$ so that the true error pattern can thus be obtained by conventional binary decoding.

Similarly, if the error pattern is not obtained when $Y' = Y \oplus T_3$, there must exist at least $(d/2) - 1$ errors outside the set of three lowest confidence values given by $S(T_3)$. Note that if two errors are located in the set $S(T_3)$, then there are only $(d/2) - 1$ errors remaining in the sequence $Y \oplus T_3$, since two errors are removed and one new error is added by forming the sequence $Y \oplus T_3$. If three errors are located in the set $S(T_3)$, then there exist only $(d/2) - 2$ errors in the sequence $Y \oplus T_3$. Thus, at most, one error can be in the set $S(T_3)$.

In general if the true error pattern is not obtained by perturbing Y with T_i, for $i = 1,3,\cdots, d-1$, there must exist at least $(d/2) - (i-1)/2$ errors outside $S(T_i)$, or equivalently, there must exist $[(d/2) - (i-1)/2]$ errors with confidence values greater than the i lowest values. This information can be represented pictorially as in Fig. 9, where all $d/2$ errors are assumed to have confidence values located in their lowest possible positions or rank.

To show that (50) is true in this case, we first combine pairwise the error with the lowest confidence value (in Fig. 9 this is an error of rank 2) with a correct digit of lower confidence value (in Fig. 9 this is one of rank 1), so that the sum of two weighted decision statistics is negative. Similarly, we can form a pairwise negative sum for each of the $d/2$ digits in error. Combining these $d/2$ negative sums yields a sum of d weighted decision statistics, which is less than zero, satisfying (50) for the case when the true error pattern has $d/2$ errors.

In the general case for any error pattern with $d/2 + i$ $(i = 1,2,\cdots, d/2 - 1)$ (50) can be shown to be true by first noting that the $d/2$ highest ranking digits in the actual error pattern must have ranks at least as high as those shown in Fig. 9. Note that if the number of errors in the actual pattern is greater than $d/2$, there must be at least (and generally more than) $d/2$ errors outside $S(T_1)$, $(d/2) - 1$ errors outside $S(T_3)$, etc. The remaining i digits of the error pattern, i.e., those that have a lower rank than the $d/2$ highest ones, can occupy at most i positions of order less than d. Thus, in Fig. 9 at most i of the available $d/2$ positions of rank less than d with $w_i v_i > 0$ (indicated by 0) can be occupied by actual errors (indicated by X). However, the required $d/2$ pairwise negative sums can still be formed by, in some cases, combining two negative values of $w_i v_i$.

We now have shown that (50) holds when the true error pattern is not selected by Algorithm 3 for error patterns of up to $d-1$ members. To complete the proof, we note that if the true error pattern contains more than $d-1$ members, the required negative sum of d terms can be formed by using d members of the actual error pattern.

B. Proof of Theorem 2

In this section we will derive the upper bound given in Theorem 2, which is valid when a threshold is used for correlation decoding and Algorithm 1-3. The set of error patterns to be tested is now required to have an analog weight less than or equal to a threshold based on the sum of the k lowest confidence values. The analog weight of this threshold is actually the analog weight of a test pattern T_k, which has 1's in the k lowest channel measurement positions and is denoted as $W_\alpha(T_k)$.

To simplify the proof of this theorem, much of the argument presented will be based on the proof of Theorem 1. We start by listing the following three events for a block error.

i) The true error pattern is considered and its analog weight is higher than some other error pattern considered. A block error must be made under this condition whether or not the analog weight of the error patterns under consideration are less than the threshold.

ii) The true error pattern is not considered by the given algorithm. (This condition cannot happen for correlation decoding.) As in the above case, a block error must be made under this condition regardless of the relative value of the threshold.

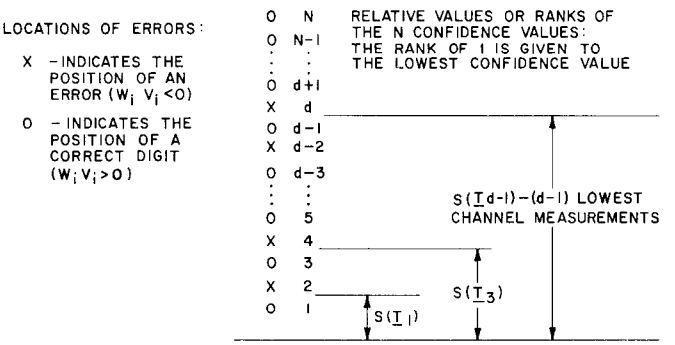

Fig. 9. Relative positions of an error pattern with $d/2$ errors, which is not tested by Algorithm 3.

iii) The true error pattern is considered by the given decoding algorithm, but cannot be selected because its analog weight is greater than the chosen threshold.

Since events i) and ii) do not depend on a threshold, we have from the proof of Theorem 1 that (50) is a necessary condition for those events to be true. Our efforts can now be devoted to finding a necessary condition for an error when event iii) occurs.

For this case we have that the analog weight of the true error pattern $W_\alpha(Z)$ is greater than $W_\alpha(T_k)$. Letting $S(T_k)$ be the set of i giving the locations of the k lowest channel measurements and $S(Z)$ the set of i for the true error pattern, we can write event iii) as

$$\sum_{i \in S(T_k)} |w_i v_i| < \sum_{i \in S(Z)} |w_i v_i|. \tag{56}$$

Assume for the moment that the number of errors in the true error pattern is r and $r \leq k$, the number of members in the set $S(T_k)$. Since (56) holds, at most $r-1$ of the $w_i v_i$ contained in $S(Z)$ can also be contained in $S(T_k)$. Thus, there exist at least $k - (r-1)$ $w_i v_i$ contained in the set $S(T_k)$ and not contained in the set $S(Z)$. Combining $k - (r-1)$ positive $w_i v_i$ contained in $S(T_k) \cap (\bar{S}Z)$ and the r negative $w_i v_i$ contained in $S(Z)$ yields a negative sum

$$\sum_{i \in S_{k+1}} w_i v_i < 0, \tag{57}$$

where S_{k+1} is a set of $k+1$ weighted decision statistics.

For the case when the number of errors in the true error pattern is greater than k, i.e., $r \geq k+1$, (57) still holds by forming a negative sum with $k+1$ members of the true error pattern.

Comparing (57) to (50), we note that the summation with the minimum number of $w_i v_i$ yields a necessary conditions for a block error when decoding with a threshold. Note that the summation with the larger number of terms is still a negative sum when some of its highest value terms are removed. Thus, an upper bound for correlation decoding and Algorithms 1-3 when a threshold of $W_\alpha(T_k)$ is used follows from (50) to (52) and is given by

$$P_e(\text{block}) < \binom{N}{d} \Pr\left[\sum_{i=1}^{d} w_i v_i < 0\right], \qquad k \geq d-1$$

$$P_e(\text{block}) < \binom{N}{k+1} \Pr\left[\sum_{i=1}^{k+1} w_i v_i < 0\right], \qquad 0 \leq k \leq d-1. \tag{58}$$

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, July/Oct., 1948, pp. 379–423.
[2] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
[3] W. W. Peterson, *Error-Correcting Codes*. Cambridge, Mass.: M.I.T. Press, and New York: Wiley, 1961.

[4] T. Kasami, "A decoding procedure for multiple-error-correcting cyclic codes," *IEEE Trans. Inform. Theory*, vol. IT-10, Apr. 1964, pp. 134–138.
[5] E. R. Berlekamp, *Algebraic Coding Theory*. New York: McGraw-Hill, 1968.
[6] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*. New York: Wiley, 1965, pp. 212–405.
[7] R. A. Silverman and M. Balser, "Coding for constant-data-rate systems—Part I. A new error-correcting code," *Proc. IRE*, vol. 42, Sept. 1954, pp. 1428–1435.
[8] M. Balser and R. A. Silverman, "Coding for constant-data-rate systems—Part II. Multiple-error-correcting codes," *Proc. IRE*, vol. 43, June 1955, pp. 728–733.
[9] C. M. Hackett, Jr., "Word error rate for group codes detected by correlation and other means," *IEEE Trans. Inform. Theory*, vol. IT-9, Jan. 1963, pp. 24–33.
[10] J. L. Massey, *Threshold Decoding*. Cambridge, Mass.: M.I.T. Press, 1963, ch. 7.
[11] G. D. Forney, Jr., "Generalized minimum distance decoding," *IEEE Trans. Inform. Theory*, vol. IT-12, Apr. 1966, pp. 125–131; also *Concatenated Codes*. Cambridge, Mass.: M.I.T. Press, 1966, ch. 3.
[12] C. R. Cahn, "Binary decoding extended to nonbinary demodulation of phase-shift keying," *IEEE Trans. Commun. Technol.* (Concise Papers), vol. COM-17, Oct. 1969, pp. 583–588.
[13] E. J. Weldon, Jr., "Encoding and decoding for binary input, M-ary output channels," presented at the IEEE Int. Symp. Information Theory, Noordwijk, the Netherlands, June 1970.
[14] W. B. Dorsch, "Maximum likelihood decoding of binary group codes for the Gaussian channel," presented at the IEEE Int. Symp. Information Theory, Noordwijk, the Netherlands, June 1970.
[15] D. Chase and R. C. Harper, "An investigation of coding and diversity techniques for a selective fading channel," presented at the IEEE Int. Conf. Communications, Boulder, Colo., June 1969.
[16] D. Chase and M. M. Goutmann, "Adaptive coding techniques for signaling over dispersive channels," presented at the URSI Spring Meeting, Washington, D.C., Apr. 1970.
[17] G. L. Turin, "Communication through noisy, random-multipath channels," *IRE Conv. Rec.*, pt. 4, Mar. 1956, pp. 154–166.
[18] W. C. Lindsey, "Error probabilities for Rician fading multichannel reception of binary and N-ary signals," *IEEE Trans. Inform. Theory*, vol. IT-10, Oct. 1964, pp. 339–350.
[19] R. Price, "Error probabilities for adaptive multichannel reception of binary signals," *IRE Trans. Inform. Theory*, vol. IT-8, Sept. 1962, pp. 305–316.
[20] G. L. Turin, "Error probabilities for binary symmetric ideal reception through nonselective slow fading and noise," *Proc. IRE*, vol. 46, Sept. 1958, pp. 1603–1619.
[21] P. A. Bello and B. D. Nelin, "Predetection diversity combining with selectively fading channels," *IRE Trans. Commun. Syst.*, vol. CS-10, Mar. 1962, pp. 32–42.
[22] ——, "The influence of the fading spectrum on the binary error probabilities of incoherent and differentially coherent matched filter receivers," *IRE Trans. Commun. Syst.*, vol. CS-10, June 1962, pp. 160–168.
[23] J. G. Proakis, "Probabilities of error for adaptive reception of M-phase signals," *IEEE Trans. Commun. Technol.*, vol. COM-16, Feb. 1968, pp. 71–81.
[24] W. C. Lindsey, "Asymptotic performance characteristics for the adaptive coherent multireceiver and noncoherent multireceiver operating through the Rician fading multichannel," *IEEE Trans. Commun. Electron.*, vol. 84, Jan. 1965, pp. 64–73; also Jet Propul. Lab., Pasadena, Calif., Rep. TR 32-440, Apr. 1963.

Random Error and Burst Correction by Iterated Codes

SUDHAKAR M. REDDY, MEMBER, IEEE, AND JOHN P. ROBINSON, MEMBER, IEEE

Abstract—**We give a decoding algorithm for iterated codes that can correct up to the number of errors guaranteed by the product minimum distance, rather than about half that number when the iterated codes are decoded independently. This result is achieved by adapting Forney's generalized minimum distance decoding for use with iterated codes. We derive results on the simultaneous burst- and random-error-correction capability of iterated codes that improve considerably on known results.**

I. Introduction

In an earlier paper, a decoding algorithm [1] that enabled correction of random errors up to the number guaranteed by the minimum distance was given for iterated codes when one of the component codes was majority decodable [2]. In the present paper, the decoding algorithm is generalized to the case where neither of the component codes is majority decodable and also to the case when the component codes can correct simultaneous burst and random errors [3]. The extensions to the case when the component codes can correct burst and random errors simultaneously considerably improve the known capabilities of iterated codes. The algorithm is an adaptation of a decoding scheme proposed by Forney for concatenated codes [4].

To minimize the introduction of new notation and also for the sake of completeness, some results of Forney are presented [4]. We treat the binary case only, though all the results are extendable in a straightforward way to the nonbinary case. Notation due to Forney [4] is used in this section.

Vectors are considered as binary vectors and also as vectors over real numbers. When vectors are considered as real vectors, binary zero is replaced by -1 and the binary 1 is considered as 1 over the real numbers. If α and x are two real vectors, then $\alpha \cdot x$ is the inner product of α and x.

The following two theorems are used to help prove the properties of the new decoding algorithm.

Theorem 1 (Forney) [4]: There is at most one codeword x_m from a binary block code of length n and minimum distance d such that $\alpha \cdot x_m > n - d$ when the coordinates

Manuscript received January 16, 1970; revised July 8, 1970; November 12, 1970; and June 5, 1971. This research was supported in part by NSF Grants GK-10025 and GK-2990.

The authors are with the Department of Electrical Engineering, University of Iowa, Iowa City, Iowa 52240.