Performance of Binary Block Codes at Low Signal-to-Noise Ratios

Chi-Chao Chao, *Member, IEEE*, Robert J. McEliece, *Fellow, IEEE*, Laif Swanson, *Member, IEEE*, and Eugene R. Rodemich

***Abstract*—The performance of general binary block codes on an unquantized additive white Gaussian noise (AWGN) channel at low signal-to-noise ratios is considered. Expressions are derived for both the block-error and the bit-error probabilities near the point where the bit signal-to-noise ratio is zero. These expressions depend on the global geometric structure of the code, although the minimum distance still seems to play a crucial role. Examples of codes such as orthogonal codes, biorthogonal codes, the (24,12) extended Golay code, and the (15,6) expurgated BCH code are discussed. The asymptotic coding gain at low signal-to-noise ratios is also studied.**

***Index Terms*—Block codes, additive white Gaussian noise channel, low signal-to-noise ratios, error probabilities, coding gain.**

I. Introduction

It is well known that for binary block codes of a fixed rate using phase-shift keying modulation on an AWGN channel, for high signal-to-noise ratios, the decoder error probability is asymptotically controlled by the code's minimum distance: the higher the minimum distance, the better the code will perform. However, for some applications, for example when the code in question is the inner code in a concatenated coding system, it is important to know about decoder error probability at *low* signal-to-noise ratios. There is, however, surprisingly little published work on this problem. In [1], Posner studied the behavior of binary block codes on an AWGN channel at low signal-to-noise ratios, but most of his results are for "hard-decision" decoders, the only exception being his "soft-decision" results for orthogonal codes. In this paper, which should be considered as a belated continuation of [1], we will study the decoder error probabilities for general binary block codes on an AWGN channel assuming maximum-likelihood, i.e., soft-decision, decoding.

Our approach is to study error probabilities of codes near the point where the bit signal-to-noise ratio E_b/N_0 is zero. One of the results obtained is the following approximation to the *block*-error probability:

$$P_E \sim 1 - \frac{1}{M} - \sqrt{\frac{E_b}{N_0}} \cdot \sqrt{\frac{R}{\pi}} \sum_{i=1}^{M-1} \sqrt{d_i} P_i, \qquad (1)$$

where the binary block code considered has M codewords and rate R, and d_i is the Hamming distance between the ith codeword $\boldsymbol{x}_i$ and the transmitted codeword $\boldsymbol{x}_0$. We shall see that the numbers P_i in (1) depend in a complicated way on the global geometric structure of the code, so that it is apparently not possible to extract from (1) a simply-interpreted quantity that controls the code's error probability at low signal-to-noise ratios. Still, we shall see in several special cases, and conjecture in general, that the terms in (1) corresponding to codewords at minimum distance from $\boldsymbol{x}_0$ dominate the expression, so that the minimum distance may still play a crucial role, even at low signal-to-noise ratios.

The full derivation of (1) is given in Section II. In Section III, a similar expression is found for the *bit-error* probability at low signal-to-noise ratios. The properties of P_i are further explored in Section IV. Then the results are applied to examples such as orthogonal codes, biorthogonal codes, the (24,12) extended Golay code and the (15,6) expurgated BCH code in Section V. In Section VI, we study the asymptotic coding gain at low signal-to-noise ratios. Finally, discussions and conjectures are given in Section VII.

II. Block-Error Probability

Let $C = \{\boldsymbol{x}_0, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_{M-1}\}$ be a binary block code (with components 0 and 1) of length n and rate $R = (\log_2 M)/n$. We will evaluate the performance of C on an unquantized AWGN channel as a function of the bit signal-to-noise ratio E_b/N_0, which we denote by λ^2. Suppose each codeword is equally likely to be selected for transmission. The codes we are interested in are all "sym-

Manuscript received November 12, 1990; revised February 25, 1992. The contributions of L. Swanson and E. Rodemich, and part of the contribution of R. McEliece, were done at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. The contribution of C. Chao while he was a doctoral student at Caltech, and part of the contribution of R. McEliece, were supported by the Air Force Office of Scientific Research under Grant No. AFOSR-91-0037. This work was presented in part at the 1988 IEEE International Symposium on Information Theory, Kobe, Japan, June 1988 and in part at the 1990 IEEE International Symposium on Information Theory, San Diego, California, Jan. 14–19, 1990.

C.-C. Chao is with the Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan 30043, R.O.C.

R. J. McEliece is with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125.

L. Swanson and E. R. Rodemich are with the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109.

IEEE Log Number 9200886.

0018-9448/92$03.00

metric" in the sense that the error probabilities are independent of which codeword is transmitted (all linear codes have this property, for example). Therefore, we assume that $\boldsymbol{x}_0$ is transmitted.

If $\hat{\boldsymbol{x}}_0$ is the counterpart of $\boldsymbol{x}_0$ with its components 0 changed to -1, then the output of the channel becomes

$$\boldsymbol{y} = \sqrt{S}\,\hat{\boldsymbol{x}}_0 + \boldsymbol{z},$$

where the quantity $\sqrt{S} = \lambda\sqrt{2R}$ and the vector $\boldsymbol{z} = (z_1, z_2, \cdots, z_n)$ has all components i.i.d. standard normal random variables (normal random variables with mean zero and variance 1). The maximum-likelihood decoder outputs the codeword with the minimum Euclidean distance from the received vector $\boldsymbol{y}$. This will be the correct decision, if and only if the decoded codeword was actually transmitted, or equivalently,

$$|\boldsymbol{z}|^2 < |\boldsymbol{y} - \sqrt{S}\,\hat{\boldsymbol{x}}_i|^2, \qquad \text{for } i = 1, 2, \cdots, M-1.$$

This inequality can be rewritten as

$$\langle \boldsymbol{z}, \sqrt{S}\,(\hat{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}_0)\rangle < \frac{1}{2}\left|\sqrt{S}\,(\hat{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}_0)\right|^2,$$

where $\langle \boldsymbol{z}, \sqrt{S}\,(\hat{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}_0)\rangle$ denotes the inner product of $\boldsymbol{z}$ and $\sqrt{S}\,(\hat{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}_0)$. Let d_i be the Hamming distance between $\boldsymbol{x}_i$ and $\boldsymbol{x}_0$ and let $\boldsymbol{u}_i$ be the vector in the direction of $\hat{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}_0$ with magnitude $\sqrt{d_i}$. (Actually $\boldsymbol{u}_i$ is just $\boldsymbol{x}_i$ if $\boldsymbol{x}_0 = \boldsymbol{0}$.) Then $\sqrt{S}\,(\hat{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}_0) = 2\lambda\sqrt{2R}\,\boldsymbol{u}_i$. If we define the normal random variables

$$T_i = \langle \boldsymbol{z}, \boldsymbol{u}_i\rangle, \qquad \text{for } i = 1, 2, \cdots, M-1, \tag{2}$$

then P_C, the probability of correct decoding, is given by

$$P_C = \Pr\left\{T_i < \lambda\sqrt{2R}\,d_i, \text{for } i = 1, 2, \cdots, M-1\right\}.$$

If the cumulative distribution function of $T_1, T_2, \cdots, T_{M-1}$ is denoted by $F(x_1, x_2, \cdots, x_{M-1})$, then P_C can be written as

$$P_C = F\left(\lambda\sqrt{2R}\,d_1, \lambda\sqrt{2R}\,d_2, \cdots, \lambda\sqrt{2R}\,d_{M-1}\right). \tag{3}$$

Note that $T_1, T_2, \cdots, T_{M-1}$ are $M-1$ normal random variables with mean zero and covariances

$$\sigma_{ij} = \langle \boldsymbol{u}_i, \boldsymbol{u}_j\rangle.$$

If $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_{M-1}$ are independent, then the covariance matrix V is nonsingular and the density function of $T_1, T_2, \cdots, T_{M-1}$ is given by

$$p(x_1, x_2, \cdots, x_{M-1}) = \frac{1}{\sqrt{(2\pi)^{M-1}|V|}}\, e^{-1/2 x^T V^{-1} x}.$$

We can therefore write P_C as an $M-1$-fold integral:

$$P_C = \int_{-\infty}^{\lambda\sqrt{2R}\,d_1} \int_{-\infty}^{\lambda\sqrt{2R}\,d_2} \cdots \int_{-\infty}^{\lambda\sqrt{2R}\,d_{M-1}} p(\boldsymbol{x})\, d\boldsymbol{x}.$$

However, if $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_{M-1}$ are not independent, then V is singular and $T_1, T_2, \cdots, T_{M-1}$ are "degenerate" in the sense of [2, p. 87], and we cannot convert P_C to an integral. This is true for most practical codes because usually $M \gg n$. For example, the (24, 12) extended Golay code has $M = 4096$ and $n = 24$.

The approach we take is to view P_C in (3) as a function of λ and approximate P_C by $P_C^{(0)} + \lambda P_C^{(1)}$, the first two terms in a power series expansion of $P_C(\lambda)$, in the neighborhood of $\lambda = 0$. The following two theorems are essential to our derivation. Their proofs are given in Appendix A.

Theorem 1: Let $X_1, X_2, \cdots, X_{M-1}$ be $M-1$ mean zero jointly normal random variables (possibly degenerate),with covariances σ_{ij} and with cumulative distribution function $F(x_1, x_2, \cdots, x_{M-1})$, with every pair X_i, X_j linearly independent (nondegenerate). Let $a_1, a_2, \cdots, a_{M-1}$ be nonnegative real numbers. Then, for $x > 0$,

$$F(a_1x, a_2x, \cdots, a_{M-1}x) = F(0, 0, \cdots, 0) + \frac{x}{\sqrt{2\pi}} \sum_{i=1}^{M-1} \frac{a_i}{\sqrt{\sigma_{ii}}} P_i + O(x^2),$$

where

$$P_i = \Pr\left\{X_j \le 0 \text{ for } j \ne i \mid X_i = 0\right\} \overset{\text{def}}{=} \lim_{h \to 0} \frac{\Pr\{X_j \le 0 \text{ for } j \ne i, 0 < X_i \le h\}}{\Pr\{0 < X_i \le h\}}.$$

The next theorem establishes that the limit P_i previously defined exists, and indeed establishes the rate at which the limit is approached.

Theorem 2: Define, for $h \ge 0$,

$$P_i(h) = \Pr\left\{X_j \le 0, \text{for } j \ne i \,|\, 0 < X_i \le h\right\}.$$

Also define the random variables Y_j, for $j \ne i$, by

$$Y_j = \sigma_{ii} X_j - \sigma_{ij} X_i.$$

Then,

$$\lim_{h \to 0} P_i(h) = P_i = \Pr\{Y_j \le 0, \text{for } j \ne i\}.$$

Also,

$$|P_i(h) - P_i| = O(h).$$

Since no two T_i's are linearly independent because no two $\boldsymbol{u}_i$'s are, we can now use Theorems 1 and 2 to estimate P_C. Note that since the codes we consider are "symmetric,"

$$F(0, 0, \cdots, 0) = \frac{1}{M}, \tag{4}$$

which follows from the fact that each codeword is equally likely to be decoded if there is no signal at all. Also note that $\sigma_{ii} = E(T_i^2) = \langle \boldsymbol{u}_i, \boldsymbol{u}_i\rangle = d_i$. Hence, we have (combining (3) and Theorem 1)

$$P_C = \frac{1}{M} + \lambda \cdot \sqrt{\frac{R}{\pi}} \sum_{i=1}^{M-1} \sqrt{d_i}\, P_i + O(\lambda^2), \tag{5}$$

where P_i is the conditional probability that $T_1 \le 0, \cdots, T_{i-1} \le 0$, $T_{i+1} \le 0, \cdots$, $T_{M-1} \le 0$, given that $T_i = 0$.

The block-error probability $P_E = 1 - P_C$, and hence, for very small λ,

$$P_E \sim 1 - \frac{1}{M} - \lambda \cdot \sqrt{\frac{R}{\pi}} \sum_{i=1}^{M-1} \sqrt{d_i} P_i,$$

which is the result of (1).

III. Bit-Error Probability

Maximum-likelihood decoding of binary *linear* block codes on an unquantized AWGN channel is now considered. We define P_b, the bit-error probability, to be the ratio of the expected number of information bits in error to the number of information bits. Let $C = \{\boldsymbol{x}_0, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_{M-1}\}$ be a binary linear block code of length n and rate $R = k/n$, where $M = 2^k$. Assume $\boldsymbol{x}_0 = \boldsymbol{0}$ is transmitted, and then

$$\boldsymbol{y} = \sqrt{S}\hat{\boldsymbol{x}}_0 + \boldsymbol{z}$$

is the received vector. If the decoder chooses to output $\boldsymbol{x}_i$, then it will make w_i bit errors, where w_i is the number of 1's in the information sequence corresponding to $\boldsymbol{x}_i$. The expected number of information bits in error is hence

$$b = \sum_{i=1}^{M-1} w_i \Pr\{\text{The decoder outputs } \boldsymbol{x}_i.\},$$

and the bit-error probability is $P_b = b/k$.

It remains to find the probability that the output codeword is $\boldsymbol{x}_i$. The maximum-likelihood decoder will output $\boldsymbol{x}_i$, if and only if the Euclidean distance between the received vector $\boldsymbol{y}$ and $\hat{\boldsymbol{x}}_i$ is the smallest among all codewords, i.e.,

$$|\boldsymbol{y} - \sqrt{S}\hat{\boldsymbol{x}}_i| < |\boldsymbol{y} - \sqrt{S}\hat{\boldsymbol{x}}_j|, \qquad \text{for } j \neq i,$$

which is equivalent to

$$\langle \boldsymbol{z}, \hat{\boldsymbol{x}}_j - \hat{\boldsymbol{x}}_i \rangle < \sqrt{S} \langle \hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}_j \rangle, \qquad \text{for } j \neq i. \quad (6)$$

If we define d_{ij} to be the Hamming distance between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ and $\boldsymbol{u}_{ij}$ to be the vector in the direction of $\hat{\boldsymbol{x}}_j - \hat{\boldsymbol{x}}_i$ with magnitude $\sqrt{d_{ij}}$, then $\langle \boldsymbol{z}, \hat{\boldsymbol{x}}_j - \hat{\boldsymbol{x}}_i \rangle = 2\langle \boldsymbol{z}, \boldsymbol{u}_{ij} \rangle$. Also $\sqrt{S}\langle \hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}_j \rangle = 2\lambda\sqrt{2R}(d_j - d_i)$. Hence, (6) is equivalent to

$$\langle \boldsymbol{z}, \boldsymbol{u}_{ij} \rangle < \lambda\sqrt{2R}(d_j - d_i), \qquad \text{for } j \neq i.$$

For $i \neq j$, we define the normal random variables $T_{ij} = \langle \boldsymbol{z}, \boldsymbol{u}_{ij} \rangle$, which have mean zero and variances d_{ij}. Then, the bit-error probability is

$$P_b = \frac{1}{k} \sum_{i=1}^{M-1} w_i \Pr\left\{T_{ij} < \lambda\sqrt{2R}(d_j - d_i), \text{for } j \neq i\right\}.$$

Now we view P_b as a function of λ and approximate P_b by $P_b^{(0)} + \lambda P_b^{(1)}$, the first two terms in a power series expansion of $P_b(\lambda)$ near $\lambda = 0$. Proceeding as in the last section, we obtain

$$P_b^{(0)} = \frac{1}{k} \sum_{i=1}^{M-1} w_i F^i(0, 0, \cdots, 0),$$

where F^i is the cumulative distribution function of $T_{i0}, \cdots, T_{i,i-1}, T_{i,i+1}, \cdots, T_{i,M-1}$. Also,

$$P_b^{(1)} = \frac{1}{k}\sqrt{\frac{R}{\pi}} \sum_{i=1}^{M\text{-}1} w_i \sum_{j \neq i} \frac{(d_j - d_i)}{\sqrt{d_{ij}}} P_{ij},$$

where P_{ij} $(i \neq j)$ is the conditional probability that $T_{ij'} \leq 0$, for $j' \neq i$ and $j' \neq j$, given that $T_{ij} = 0$. The linearity of the code is now used to simplify both the expressions of $P_b^{(0)}$ and $P_b^{(1)}$. For every pair of codewords $\boldsymbol{x}_i, \boldsymbol{x}_j$, there always exists another codeword $\boldsymbol{x}_l$ such that $\boldsymbol{x}_i \oplus \boldsymbol{x}_j = \boldsymbol{x}_l$, where $\oplus$ is the modulo-2 addition. Since the normal distribution is symmetric about the origin, up to a permutation of the parameters, F^i, for $i = 1, 2, \cdots, M-1$ are equivalent to F in the last section, and $P_{ij} = P_l$, where $\boldsymbol{x}_l = \boldsymbol{x}_i \oplus \boldsymbol{x}_j$. By (4), it follows that

$$P_b^{(0)} = \frac{1}{kM} \sum_{i=1}^{M-1} w_i = \frac{1}{kM} \cdot \frac{Mk}{2} = \frac{1}{2},$$

as expected, since each information bit is correct with probability 1/2 when there is no signal. Now using the fact that $d_{ij} = d_l$ if $\boldsymbol{x}_i \oplus \boldsymbol{x}_j = \boldsymbol{x}_l$, we obtain

$$P_b \sim \frac{1}{2} - \lambda \cdot \frac{1}{k}\sqrt{\frac{R}{\pi}} \sum_{i=1}^{M-1} w_i \sum_{\substack{j \neq i \\ \boldsymbol{x}_i \oplus \boldsymbol{x}_j = \boldsymbol{x}_l}} \frac{(d_i - d_j)}{\sqrt{d_l}} P_l. \quad (7)$$

The previous approximation applies to all binary linear block codes. Further simplifications can be obtained if more assumptions are made. Now suppose C is systematic and has the symmetry property such that each bit in the codeword is "permutationally equivalent" to each other bit, e.g., C is cyclic, or more generally, its automorphism group (see definition in Section IV) contains a transitive permutation group. Then the bit-error probability can be found, alternatively, by dividing the expected number of codeword bits in error by the block length n. All the derivations remain the same as before except that k and w_i will now be replaced by n and d_i, respectively. Then,

$$P_b \sim \frac{1}{2} - \lambda \cdot \frac{1}{n}\sqrt{\frac{R}{\pi}} \sum_{i=1}^{M-1} d_i \sum_{\substack{j \neq i \\ \boldsymbol{x}_i \oplus \boldsymbol{x}_j = \boldsymbol{x}_l}} \frac{(d_i - d_j)}{\sqrt{d_l}} P_l. \quad (8)$$

Algebraic manipulations show that

$$\sum_{i=1}^{M\text{-}1} d_i \sum_{\substack{j \neq i \\ \boldsymbol{x}_i \oplus \boldsymbol{x}_j \\ = \boldsymbol{x}_l}} \frac{(d_i - d_j)}{\sqrt{d_l}} P_l = \sum_{l=1}^{M-1} \frac{P_l}{\sqrt{d_l}} \sum_{i=0}^{M-1} d_i(d_i - d(\boldsymbol{x}_i \oplus \boldsymbol{x}_l)),$$

where $d(\boldsymbol{x}_i \oplus \boldsymbol{x}_l)$ denotes the Hamming distance between $\boldsymbol{x}_i \oplus \boldsymbol{x}_l$ and $\boldsymbol{x}_0$. If we further assume that C contains no repeated columns, i.e,. there are no two positions in the block where the corresponding bits are the same for all codewords, then from Appendix B,

$$\sum_{i=1}^{M-1} d_i^2 = n(n+1)2^{k-2}, \tag{9}$$

and

$$\sum_{i=1}^{M\text{-}1} d_i d(\boldsymbol{x}_i \oplus \boldsymbol{x}_l) = n(n+1)2^{k-2} - d_l 2^{k-1}. \tag{10}$$

Equation (8) can hence be written as

$$P_b \sim \frac{1}{2} - \lambda \cdot \frac{M}{2n}\sqrt{\frac{R}{\pi}} \sum_{i=1}^{M\text{-}1} \sqrt{d_i} P_i. \tag{11}$$

Note that (1) and (11) have similar forms, and in particular that both involve the mysterious quantities P_i, which are the subject of the next section.

IV. Properties of P_i

The probability P_i is the conditional probability that $T_1 \le 0, \cdots, T_{i-1} \le 0, T_{i+1} \le 0, \cdots, T_{M-1} \le 0$, given that $T_i = 0$. In order to illustrate the calculation of P_i, consider the $M = 4$ orthogonal code $\{\boldsymbol{x}_0 = 0000, \boldsymbol{x}_1 = 0101, \boldsymbol{x}_2 = 0011, \boldsymbol{x}_3 = 0110\}$. By (2), $T_1 = z_2 + z_4, T_2 = z_3 + z_4, T_3 = z_2 + z_3$, where z_1, z_2, z_3, z_4 are i.i.d. standard normal random variables. Therefore,

$$P_1 = \Pr\{T_2 \le 0, T_3 \le 0 \mid T_1 = 0\},$$
$$P_2 = \Pr\{T_1 \le 0, T_3 \le 0 \mid T_2 = 0\},$$
$$P_3 = \Pr\{T_1 \le 0, T_2 \le 0 \mid T_3 = 0\}.$$

Since z_2, z_3, z_4 are i.i.d., it is easy to see that $P_1 = P_2 = P_3$. It remains to find the probability that $z_3 + z_4 \le 0$, $z_2 + z_3 \le 0$, given that $z_2 + z_4 = 0$, which is the conditional probability that a random point with a normal distribution in a three-dimensional space falls in the region described by $z_3 + z_4 \le 0$, $z_2 + z_3 \le 0$ given that it is on the plane $z_2 + z_4 = 0$. We will see in the next section that $P_1 = P_2 = P_3 = \tan^{-1}\sqrt{2}/\pi$. However, for most practical codes with $M \gg n$, it appears to be very difficult (if not impossible) to obtain a closed-form expression for P_i.

Definition 1: The set of coordinate permutations that map every codeword in the code $\boldsymbol{C}$ into a (possibly different) codeword in $\boldsymbol{C}$ is called the automorphism group of $\boldsymbol{C}$, denoted by Aut$(\boldsymbol{C})$.

It is known that Aut$(\boldsymbol{C})$ is indeed a group. Automorphism groups of several block codes are discussed in [3]–[5]. There are computer search algorithms [6], [7] for finding the entire automorphism group of a code. Furthermore, the entire automorphism groups of all 2, 3, 4-error correcting binary primitive BCH codes have been determined algebraically in [8]. Note that the permutations in Aut$(\boldsymbol{C})$ partition the codewords in $\boldsymbol{C}$ into equivalence classes. Codewords $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are in the same equivalence class if there exists a permutation in Aut$(\boldsymbol{C})$ that maps $\boldsymbol{x}_i$ to $\boldsymbol{x}_j$.

Theorem 3: Assume $\boldsymbol{x}_0 = \boldsymbol{0}$. If $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are in the same equivalence class partitioned by permutations in Aut$(\boldsymbol{C})$, then $P_i = P_j$.

Proof: If $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are in the same equivalence class, a permutation ϕ that maps $\boldsymbol{x}_i$ to $\boldsymbol{x}_j$ will map all the codewords other than $\boldsymbol{x}_0$ and $\boldsymbol{x}_i$ to codewords other than $\boldsymbol{x}_0$ and $\boldsymbol{x}_j$. It is impossible that tow different codewords are mapped to the same codeword because ϕ^{-1} is also in Aut$(\boldsymbol{C})$. Then, T_i will be accordingly mapped to T_j and $\{T_l: l \ne i\}$ to $\{T_l: l \ne j\}$, which implies that $P_i = P_j$. □

In the preceding orthogonal code example, the permutation $\phi = (234)$ maps $\boldsymbol{x}_1$ to $\boldsymbol{x}_3$, $\boldsymbol{x}_3$ to $\boldsymbol{x}_2$ and $\boldsymbol{x}_2$ to $\boldsymbol{x}_1$, so $P_1 = P_2 = P_3$. For some practical codes, all the codewords of the same weight are in one equivalence class (but this is not generally true), so their corresponding P_i's are equal. Thus, the notation P_d is used for all the codewords of weight d. For this case, (1) can be simplified to

$$P_E \sim 1 - \frac{1}{M} - \lambda \cdot \sqrt{\frac{R}{\pi}} \sum_d A_d \sqrt{d} P_d, \tag{12}$$

where $A(z) = \sum_d A_d z^d$ is the weight enumerator. Similarly, we can simplify (11) to

$$P_b \sim \frac{1}{2} - \lambda \cdot \frac{M}{2n}\sqrt{\frac{R}{\pi}} \sum_d A_d \sqrt{d} P_d. \tag{13}$$

Recall that the original assumption for (11) is that $\boldsymbol{C}$ is linear systematic with no repeated columns and Aut$(\boldsymbol{C})$ contains a transitive permutation group.[1]

Definition 2: Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be binary vectors. If $\boldsymbol{u}$ has a 1 in every position that $\boldsymbol{v}$ has a 1, then we say that $\boldsymbol{u}$ covers $\boldsymbol{v}$.

Theorem 4: Assume $\boldsymbol{x}_0 = \boldsymbol{0}$. For a binary linear block code, if the codeword $\boldsymbol{x}_i$ covers a different nonzero codeword $\boldsymbol{x}_j$, then $P_i = 0$.

Proof: Let $\boldsymbol{x}_l = \boldsymbol{x}_i \oplus \boldsymbol{x}_j$. Then, $\boldsymbol{x}_l$ is covered by $\boldsymbol{x}_i$ and $T_i = T_j + T_l$. The only case that both T_j and T_l are less than or equal to 0 given that $T_i = 0$ is when $T_j = 0$ and $T_l = 0$. The theorem follows from the fact that all T_i's are continuous random variables. □

For most practical codes, $M \gg n$, which means that there are many more random variables T_j in the definition of the P_i's than the code dimension n. Hence, it is desirable to eliminate some redundant random variables T_j to reduce the complexity of computing P_i. One simple result is that the condition $T_l \le 0$ can be eliminated from P_i if the codeword $\boldsymbol{x}_l$ covers another nonzero codeword $\boldsymbol{x}_j$ with $T_j \le 0$. This is proved by letting $\boldsymbol{x}_m = \boldsymbol{x}_l \oplus \boldsymbol{x}_j$, and then the conditions $T_l = T_j + T_m$ and $T_j \le 0$, $T_m \le 0$ guarantee that $T_l \le 0$. The next theorem, which is proved in Appendix C by using the Farkas Alternative [9, p. 56] tells us in general how we can eliminate redundant T_j's. In

[1] A permutation group G is transitive if, for any two symbols i and j, there is a permutation $\phi \in G$ such that $i\phi = j$.

the theorem, $\boldsymbol{A}$ is a matrix and $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{d}, \boldsymbol{b}$ are column vectors, and we say a vector $\boldsymbol{x} \leq 0$ if all of its components ≤ 0.

Theorem 5: Let the set $\mathscr{A} = \{\boldsymbol{x}\colon \boldsymbol{AX} \leq 0 \text{ and } \boldsymbol{d}^T\boldsymbol{x} = 0\}$ be nonempty. The inequality $\boldsymbol{b}^T\boldsymbol{x} \leq 0$ holds for all $\boldsymbol{x} \in \mathscr{A}$, if and only if $\boldsymbol{b} = \boldsymbol{A}^T\boldsymbol{y} + \alpha\boldsymbol{d}$, for some $\boldsymbol{y} \geq 0$ and $\alpha \in R$.

To interpret this theorem, we view each $T_i \leq 0$ as an inequality in $z_1, z_2, \cdots, z_n$. The theorem implies that given $T_j \leq 0$, for $j \neq i$, and $T_i = 0$, the particular $T_l \leq 0$ is redundant and can be eliminated, if and only if $T_l = \sum_{j \neq i, l} a_j T_j + \alpha T_i$, where $a_j \geq 0$ and $\alpha \in R$. Note that setting $\alpha = 0$ reduces to the case stated previously. On the other hand, if we somehow want to create another redundant inequality $T_l \leq 0$, then T_l must be in the form of $\sum_{j \neq i}^{M-1} a_j T_j + \alpha T_i$ with $a_j \geq 0$ and $\alpha \in R$.

Note that $\sigma_{ii} = \langle \boldsymbol{u}_i, \boldsymbol{u}_i \rangle = d_i$, and for codes with $\boldsymbol{x}_0 = \boldsymbol{0}$, σ_{ij} is the number of positions where $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are both 1. As previously mentioned, for most practical codes of interest, $M \gg n$; even after the redundant $T_j < 0$ are eliminated by Theorem 5, the number of remaining conditions is still very large compared with the code dimension n. Hence, it may be necessary to resort to Monte Carlo simulations to find approximate values for the P_i. Conditional probabilities are relatively difficult to work with, but Theorem 2 gives an alternate, unconditional, formula for the P_i's, which lends itself more easily to simulation: $P_i = \Pr\{\sigma_{ii}T_j - \sigma_{ij}T_i \leq 0, \text{ for } j \neq i\}$. First n i.i.d. standard normal random variables z_i, for $i = 1, 2, \cdots, n$, are generated; then all necessary (nonredundant) conditions $\sigma_{ii}T_j - \sigma_{ij}T_i \leq 0$ are tested. If all are satisfied, we record this event as a "success." If any one of the conditions fails, we record this event as a "failure." The procedure is repeated a large number of times; then the relative frequency of "success" will be an approximate value for P_i.

V. Examples

We now apply the results in previous sections to orthogonal codes, biorthogonal codes, the (24, 12) extended Golay code, and the (15, 6) expurgated BCH code.

A. Orthogonal Codes

We consider orthogonal codes with $M = 2^k$ codewords, which are obtained from rows of $2^k \times 2^k$ normalized Hadamard matrices via the mapping that the $+1$'s are changed to 0's and the -1's to 1's. Such Hadamard matrices can be constructed by the Sylvester method [3, p. 45]. All the nonzero codewords have weights 2^{k-1} and it is easy to see that they are in the same equivalence class. By (12) near $\lambda = 0$, the block-error probability can be approximated by

$$P_E \sim 1 - \frac{1}{2^k} - \lambda \cdot (2^k - 1)\sqrt{\frac{k}{2\pi}}\, P_{2^{k-1}}, \tag{14}$$

where $P_{2^{k-1}}$ is for codewords of weight 2^{k-1}. By using $P_b = (2^{k-1}P_E)/(2^k - 1)$ [10, p. 100] or (7), then

$$P_b \sim \frac{1}{2} - \lambda \cdot 2^{k-1}\sqrt{\frac{k}{2\pi}}\, P_{2^{k-1}}. \tag{15}$$

We now want to compute the value of $P_{2^{k-1}}$, which is the conditional probability that $T_1 \leq 0, T_2 \leq 0, \cdots, T_{2^k-2} \leq 0$, given that $T_{2^k-1} = 0$. By the structure of orthogonal codes, T_i, for $i = 1, 2, \cdots, 2^k - 1$, are normal random variables with mean zero and covariances

$$\sigma_{ij} = \begin{cases} 2^{k-1}, & \text{if } i = j; \\ 2^{k-2}, & \text{if } i \neq j. \end{cases}$$

An easy calculation verifies that the random variable T_i, for $i = 1, 2, \cdots, 2^k - 1$, can be modeled by

$$T_i = \sqrt{2^{k-2}}\,(X_i + X_0),$$

where $X_0, X_1, \cdots, X_{2^k-1}$ are i.i.d. standard normal random variables. Hence,

$$\begin{aligned} P_{2^{k-1}} &= \lim_{h \to 0} \Pr\{X_1 + X_0 \leq 0, \cdots, X_{2^k-2} + X_0 \leq 0 \,|\, 0 < X_{2^k-1} + X_0 \leq h\} \\ &= \lim_{h \to 0} \frac{\Pr\{X_1 \leq -X_0, \cdots, X_{2^k-2} \leq -X_0, -X_0 < X_{2^k\text{-}1} \leq -X_0 + h\}}{\Pr\{0 < X_{2^k-1} + X_0 \leq h\}}. \end{aligned}$$

Since $X_{2^k-1} + X_0$ is normal with mean zero and variance 2, $\Pr\{0 < X_{2^k-1} + X_0 \leq h\} = h/\sqrt{4\pi} + O(h^2)$ as $h \to 0$. We also have

$$\begin{aligned} &\lim_{h \to 0} \Pr\{X_1 \leq -X_0, \cdots, X_{2^k-2} \leq -X_0, -X_0 \\ &\qquad < X_{2^k-1} \leq -X_0 + h\} \\ &\quad = \lim_{h \to 0} \int_{-\infty}^{\infty} h \cdot Z(-t)[P(-t)]^{2^k-2} Z(t)\, dt + O(h^2) \\ &\quad = \lim_{h \to 0} \frac{h}{\sqrt{2\pi}} \int_{-\infty}^{\infty} Z(\sqrt{2}t)[P(t)]^{2^k-2}\, dt + O(h^2), \end{aligned}$$

where $Z(x) = e^{-x^2/2}/\sqrt{2\pi}$ and $P(x) = \int_{-\infty}^{x} Z(t)\, dt$. Finally, we obtain

$$P_{2^{k-1}} = \sqrt{2}\int_{-\infty}^{\infty} Z(\sqrt{2}t)[P(t)]^{2^k-2}\, dt. \tag{16}$$

The same result was obtained in [1] by directly expanding into a power series the expressions of the error probabilities for orthogonal codes from [11]. Our $P_{2^{k-1}}$ is equal to $\sqrt{2}A_{2^k-1}$ in the notation of [1]. In particular, for $k = 2$, A_3 was shown to be $\tan^{-1}\sqrt{2}/(\pi\sqrt{2})$; it follows that $P_2 = \tan^{-1}\sqrt{2}/\pi$. Since it was shown that $A_v \approx$

$(2/\nu^2)\sqrt{\pi}\ln\nu$ for large ν,

$$P_{2^{k-1}} \approx \frac{2}{(2^k-1)^2}\sqrt{2\pi\ln(2^k-1)}, \qquad \text{for large } k.$$

Equation (16) has been integrated numerically for $k = 2$ to 10, and the results are listed in Table I, and so are the quantities $(2^k - 1)\sqrt{k/(2\pi)}\,\mathrm{P}_{2^{k-1}}$ and $2^{k-1}\sqrt{k/(2\pi)}\,P_{2^{k-1}}$, which are the key elements of (14) and (15), respectively. Note that, for orthogonal codes at very low signal-to-noise ratios, the bit-error probability increases with k, or the number of codewords M.

B. Biorthogonal Codes

A biorthogonal code defined here consists of the codewords of an orthogonal code and their complements. We consider biorthogonal codes with $M = 2^k$, $k \geq 2$, codewords. The biorthogonal code is the first-order Reed–Muller code if the corresponding orthogonal code is obtained by the Sylvester construction [3, p. 373]. All the codewords except the all-zero and all-one codewords have weights 2^{k-2}.

Proposition 1: All the codewords except the all-zero and all-one codewords in a biorthogonal code are in one equivalence class.

Proof: For a biorthogonal code with $M = 2^k$, $k \geq 2$, there are $2^k - 2$ codewords of weight 2^{k-2}; $2^{k-1} - 1$ of them begin with 0 because they are codewords of an orthogonal code, and the rest begin with 1. Since all the nonzero codewords in an orthogonal code are in one equivalence class, it follows that there are at most two equivalence classes for codewords of weight 2^{k-2} in a biorthogonal code, one for each half. However, since the automorphism group of a Reed–Muller code contains the general affine group that is triply transitive[2] [3, pp. 398–400], there exist permutations in the automorphism group of a biorthogonal code (which is the first-order Reed–Muller code) that map a nonzero codeword beginning with 0 to a codeword beginning with 1. The proposition hence follows. □

The all-one codeword covers every codeword of weight 2^{k-2}, so from Theorem 4 the corresponding P_i is zero. By (12), we now have

$$P_E \sim 1 - \frac{1}{2^k} - \lambda\cdot(2^k-2)\sqrt{\frac{k}{2\pi}}\,P_{2^{k-2}}.$$

TABLE I
$P_{2^{k-1}}$ for Orthogonal Codes

k	$P_{2^{k-1}}$	$(2^k - 1)\sqrt{k/(2\pi)}\,P_{2^{k-1}}$	$2^{k-1}\sqrt{k/(2\pi)}\,P_{2^{k-1}}$
2	3.0409*e*-1	5.1469*e*-1	3.4312*e*-1
3	9.0117*e*-2	4.3589*e*-1	2.4908*e*-1
4	2.6084*e*-2	3.1219*e*-1	1.6650*e*-1
5	7.3959*e*-3	2.0453*e*-1	1.0556*e*-1
6	2.0606*e*-3	1.2686*e*-1	6.4436*e*-2
7	5.6580*e*-4	7.5845*e*-2	3.8221*e*-2
8	1.5351*e*-4	4.4170*e*-2	2.2171*e*-2
9	4.1242*e*-5	2.5222*e*-2	1.2636*e*-2
10	1.0991*e*-5	1.4185*e*-2	7.0995*e*-3

TABLE II
$P_{2^{k-2}}$ FOR BIORTHOGONAL CODES

k	$P_{2^{k-2}}$	$(2^k - 2)\sqrt{k/(2\pi)}\,P_{2^{k-2}}$
3	1.0817*e*-1	4.4848*e*-1
4	2.8223*e*-2	3.1526*e*-1
5	7.6703*e*-3	2.0527*e*-1
6	2.0968*e*-3	1.2704*e*-1
7	5.7062*e*-4	7.5889*e*-2
8	1.5415*e*-4	4.4180*e*-2
9	4.1327*e*-5	2.5225*e*-2
10	1.1003*e*-5	1.4186*e*-2
11	2.9114*e*-6	7.8817*e*-3

Since a biorthogonal code contains no repeated columns, can be encoded as a systematic code, and its automorphism group contains a triply transitive group, by (13) for λ near 0,

$$P_b \sim \frac{1}{2} - \lambda\cdot(2^k-2)\sqrt{\frac{k}{2\pi}}\,P_{2^{k-2}}.$$

Now, our goal is to find an analytical expression for $P_{2^{k-2}}$, the conditional probability that $T_2 \leq 0, T_3 \leq 0,\cdots,T_{2^k-1} \leq 0$, given that $T_1 = 0$. (Here, we number the codewords in such a way that $\boldsymbol{x}_i$, for $i = 0,1,\cdots,2^{k-1}-1$, are codewords of a corresponding orthogonal code and $\boldsymbol{x}_{2^{k-1}+i}$, for $i = 0,1,\cdots,2^{k-1}-1$, are the complements of $\boldsymbol{x}_i$.) Since the all-one codeword $\boldsymbol{x}_{2^{k-1}}$ covers every codeword of weight 2^{k-2}, the condition $T_{2^{k-1}} \leq 0$ is redundant and can be discarded. From the structure of biorthogonal codes, the covariances between T_i and T_j, for $i,j = 1,\cdots,2^{k-1}-1, 2^{k-1}+1,\cdots,2^k-1$, are given by

$$\sigma_{ij} = \begin{cases} 2^{k-2}, & \text{if } i = j; \\ 0, & \text{if } |i-j| = 2^{k-1}; \\ 2^{k-3}, & \text{otherwise.} \end{cases}$$

We then model the random variables $T_1,\cdots,T_{2^{k-1}-1}$, $T_{2^{k-1}+1},\cdots,T_{2^k-1}$ by

$$T_i = \sqrt{2^{k-3}}\,(X_0 + X_i),$$

and

$$T_{2^{k-1}+i} = \sqrt{2^{k-3}}\,(X_0 - X_i),$$

where $i = 1,2,\cdots,2^{k-1}-1$ and $X_0, X_1,\cdots,X_{2^k-1}$ are i.i.d.

[2] A permutation group G is t-fold transitive if, given t distinct symbols $i_1, i_2,\cdots,i_t$, and t distinct symbols $j_1, j_2,\cdots,j_t$, there is a permutation $\phi \in G$ such that $i_1\phi = j_1$, $i_2\phi = j_2,\cdots,i_t\phi = j_t$.

standard normal random variables. Thus,

$$\begin{aligned} P_{2^k-2} &= \lim_{h\to 0} \Pr\{X_0 - X_1 \le 0 \text{ and } X_0 + X_i \le 0,\ X_0 - X_i \le 0, \text{for } i = 2,3,\cdots,2^{k-1} - |0 < X_0 + X_1 \le h\} \\ &= \lim_{h\to 0} \Pr\{X_0 \le 0,\ -X_0 < X_1 \le -X_0 + h, \text{and } X_0 \le X_i \le -X_0, \\ &\quad \text{for } i = 2,3,\cdots,2^{k-1} - 1\}/\Pr\{0 < X_0 + X_1 \le h\} \\ &= \lim_{h\to 0} \frac{\int_{-\infty}^{0} h\cdot Z(-t)[P(-t) - P(t)]^{2^{k-1}-2} Z(t)\,dt + O(h^2)}{h/\sqrt{4\pi} + O(h^2)} \\ &= \sqrt{2}\int_0^\infty Z(\sqrt{2}t)[P(t) - P(-t)]^{2^{k-1}-2}\,dt. \end{aligned}$$

The same result can be obtained if we expand into a power series in λ the expressions for error probabilities in [11]. We have integrated numerically the expressions for P_{2^k-2}, $k = 3,4,\cdots,11$, and listed the results in Table II, along with the quantities $(2^k - 2)\sqrt{k/(2\pi)}\,P_{2^k-2}$. Again note that for biorthogonal codes at very low signal-to-noise ratios the bit-error probability increases with the number of codewords.

C. The (24, 12) Extended Golay Code

The (24, 12) extended Golay code is obtained by adding an overall parity check bit to the perfect triple-error-correcting (23, 12) Golay code. Its weight enumerator is $A(z) = 1 + 759z^8 + 2576z^{12} + 759z^{16} + z^{24}$. Note that the codeword of weight 24 is the all-one codeword, which covers all other nonzero codewords. The automorphism group of the (24, 12) Golay code is the Mathieu group M_{24} which is five-fold transitive [3, pp. 636–641].

Proposition 2 [3, p. 638]: All the codewords of weight 8 are in one equivalence class.

Proposition 3 [3, p. 641]: All the codewords of weight 12 are in one equivalence class.

Proposition 4: All the codewords of weight 16 are in one equivalence class.

Proof: The permutation that maps one codeword to another codeword will do the same to their complements. Since the complement of any codeword of weight 8 is a codeword of weight 16 and vice versa, the proposition follows from Proposition 2. □

Proposition 5: Every codeword of weight 16 covers codewords of weight 8.

Proof: By Proposition 4, this proposition can be proved by finding an instance that a codeword of weight 16 covers a codeword of weight 8. This is easily done by inspecting a generator matrix for the (24, 12) extended Golay code, e.g., the one in [3, p. 65]. □

Now, near $\lambda = 0$ the block-error probability can be approximated by

$$P_E \sim \frac{4095}{4096} - \lambda\cdot\sqrt{\frac{1}{2\pi}}\left(759\sqrt{8}\,P_8 + 2576\sqrt{12}\,P_{12}\right),$$

and the bit-error probability can be approximated by

$$P_b \sim \frac{1}{2} - \lambda\cdot\frac{256}{3}\sqrt{\frac{1}{2\pi}}\left(759\sqrt{8}\,P_8 + 2576\sqrt{12}\,P_{12}\right).$$

Note that the codewords of weight 16 and 24 play no part in the approximations for P_E and P_b, neither do they in P_8 and P_{12}. Unlike the last two examples, we do not expect exact analytical expressions for P_8 and P_{12}. The procedure described in the last section is used to simulate P_8 and P_{12}. All the inequalities corresponding to the all-one codeword and codewords of weight 16 can be discarded because of Theorem 5. The results are $P_8 \approx 4.0 \times 10^{-6}$ and $P_{12} \approx 4 \times 10^{-8}$. (Since P_{12} is very small, the reliability of the value obtained is doubtful but the magnitude should be correct.) Then,

$$\begin{aligned} P_E &\sim \frac{4095}{4096} - \lambda\cdot\sqrt{\frac{1}{2\pi}}\,(8.6\times 10^{-3} + 3.6\times 10^{-4}) \\ &\approx \frac{4095}{4096} - \lambda\cdot(3.6\times 10^{-3}), \end{aligned}$$

and

$$\begin{aligned} P_b &\sim \frac{1}{2} - \lambda\cdot\frac{256}{3}\sqrt{\frac{1}{2\pi}}\,(8.6\times 10^{-3} + 3.6\times 10^{-4}) \\ &\approx \frac{1}{2} - \lambda\cdot(0.30). \end{aligned}$$

Note that the terms above for codewords of weight 8 (codewords at the minimum distance) are much larger than the terms for P_{12}, so that in a sense, the behavior of the code at low signal-to-noise ratios is controlled by the minimum distance.

D. The (15, 6) Expurgated BCH Code

We now consider the (15, 6) expurgated BCH code with generator polynomial $(x^4 + x + 1)(x^4 + x^3 + x^2 + x + 1)(x + 1) = x^9 + x^6 + x^5 + x^4 + x + 1$. Its weight enumerator is $A(z) = 1 + 30z^6 + 15z^8 + 18z^{10}$. It is known

[7], [8] that the complete automorphism group of the (15, 7) BCH code with $d_{\min} = 5$ is the group $\{bx^{2^i} + b'x^{2^{i+2}}: b, b' \in GF(2^4), b^{2^2+1} \neq b'^{2^2+1}$ and $i = 0, 1\}$. We also know that the automorphism group of an expurgated code contains that of the original code. By examining the codewords of the (15, 6) expurgated BCH code, it can then be shown that all the codewords of equal weight are in the same equivalence classes. Therefore, the error probabilities near $\lambda = 0$ are approximately

$$P_E \sim \frac{63}{64} - \lambda \cdot \sqrt{\frac{2}{5\pi}}\left(30\sqrt{6}P_6 + 15\sqrt{8}P_8 + 18\sqrt{10}P_{10}\right),$$

and

$$P_b \sim \frac{1}{2} - \lambda \cdot \frac{32}{15}\sqrt{\frac{2}{5\pi}}\left(30\sqrt{6}P_6 + 15\sqrt{8}P_8 + 18\sqrt{10}P_{10}\right).$$

From a Monte Carlo simulation, $P_6 \approx 4.1 \times 10^{-3}$, $P_8 \approx 8.9 \times 10^{-4}$ and $P_{10} \approx 9.4 \times 10^{-5}$. Thus,

$$P_E \sim \frac{63}{64} - \lambda \cdot \sqrt{\frac{2}{5\pi}}\,(0.30 + 3.8 \times 10^{-2} + 5.4 \times 10^{-3})$$
$$\approx \frac{63}{64} - \lambda \cdot (0.12),$$

and

$$P_b \sim \frac{1}{2} - \lambda \cdot (0.26).$$

Note again that the terms above for codewords at the minimum distance are much larger than the remaining terms.

VI. Asymptotic Coding Gain

The *coding gain* is the ratio of the signal-to-noise ratio without coding to the signal-to-noise ratio required when using an error-correcting code to achieve the same error probability. We define the *asymptotic coding gain* as the limit, as the signal-to-noise ratio approaches zero, of the coding gain. Two theorems based on the criteria of P_E and P_b, respectively, will be given.

We now derive approximations to P_E and P_b at low signal-to-noise ratios when no coding is used. For an unquantized AWGN channel, the bit-error probability without coding is

$$P_b = Q(\sqrt{2}\lambda),$$

where $Q(x) = \int_x^\infty e^{-t^2/2}/\sqrt{2\pi}\,dt$. Thus, near $\lambda = 0$,

$$P_b \sim \frac{1}{2} - \lambda \cdot \frac{1}{\sqrt{\pi}}. \tag{17}$$

If k bits are grouped as a block without coding, a block error occurs when there is at least one erroneous bit and so

$$P_E = 1 - (1 - P_b)^k,$$

which gives the following approximation near $\lambda = 0$:

$$P_E \sim 1 - \frac{1}{2^k} - \lambda \cdot \frac{k}{2^{k-1}\sqrt{\pi}}. \tag{18}$$

Comparing (1), (7), (11), (17), and (18), we obtain the following theorems.

Theorem 6: For binary block codes, with the criterion based on block-error probability, the asymptotic coding gain at low signal-to-noise ratios is given by

$$G_E = \frac{2^{2(k-1)}}{nk}\left(\sum_{i=1}^{M-1}\sqrt{d_i}P_i\right)^2.$$

Theorem 7: For binary linear block codes, with the criterion based on bit-error probability, the asymptotic coding gain at low signal-to-noise ratios is given by

$$G_b = \frac{1}{nk}\left(\sum_{i=1}^{M-1} w_i \sum_{\substack{j \neq i \\ x_i \oplus x_j = x_l}} \frac{(d_i - d_j)}{\sqrt{d_l}}P_l\right)^2.$$

If the code used is systematic with no repeated columns and its automorphism group contains a transitive permutation group, then the asymptotic coding gain can be simplified to

$$G_b = \frac{k}{n}\left(\frac{2^{k-1}}{n}\sum_{i=1}^{M-1}\sqrt{d_i}P_i\right)^2,$$

which is equal to $(k/n)^2 G_E$.

We now apply the results in Theorems 6 and 7 to the codes discussed in Section V. For orthogonal codes with 2^k codewords, based on the P_E-criterion, the asymptotic coding gain is

$$G_E = \frac{2^{2(k-1)}(2^k - 1)^2}{k}\left(\int_{-\infty}^{\infty} Z(\sqrt{2}t)[P(t)]^{2^k-2}\,dt\right)^2,$$

which approaches $\pi \ln 2 \approx 3.38$ dB as $k \to \infty$. For the P_b-criterion, the asymptotic coding gain becomes

$$G_b = k2^{2(k-1)}\left(\int_{-\infty}^{\infty} Z(\sqrt{2}t)[P(t)]^{2^k-2}\,dt\right)^2,$$

which is asymptotic in k to $(\pi \ln 2)k^2/2^{2k}$. The same results were obtained in [1]. The asymptotic coding gains based on criterions P_E and P_b for orthogonal codes are listed in Table III. Note that, based on the P_E-criterion, except for $k = 2$, orthogonal codes result in positive coding gain compared with no coding at low signal-to-noise ratios and the gain increases with the number of codewords. However, for the P_b-criterion, there is always a coding loss when using an orthogonal code and the loss increases with k.

For biorthogonal codes with 2^k codewords, based on the P_E-criterion,

$$G_E = \frac{2^{2(k-1)}(2^k - 2)^2}{k} \cdot \left(\int_0^{\infty} Z(\sqrt{2}t)[P(t) - P(-t)]^{2^{k-1}-2}\,dt\right)^2.$$

TABLE III
ASYMPTOTIC CODING GAIN FOR ORTHOGONAL CODES

k	G_E	G_b
2	−0.798 dB	−4.32 dB
3	0.258 dB	−7.10 dB
4	0.880 dB	−10.6 dB
5	1.29 dB	−14.6 dB
6	1.58 dB	−18.8 dB
7	1.79 dB	−23.4 dB
8	1.96 dB	−28.1 dB
9	2.09 dB	−33.0 dB
10	2.19 dB	−38.0 dB

TABLE IV
ASYMPTOTIC CODING GAIN FOR BIORTHOGONAL CODES

k	G_E	G_b
3	0.505 dB	−1.99 dB
4	0.965 dB	−5.06 dB
5	1.32 dB	−8.78 dB
6	1.59 dB	−12.9 dB
7	1.80 dB	−17.4 dB
8	1.96 dB	−22.1 dB
9	2.09 dB	−27.0 dB
10	2.19 dB	−32.0 dB
11	2.28 dB	−37.1 dB

For the P_b-criterion,

$$G_b = k(2^k - 2)^2 \left(\int_0^{\infty} Z(\sqrt{2}t)[P(t) - P(-t)]^{2^{k-1}-2}\, dt \right)^2.$$

The asymptotic coding gains for biorthogonal codes are tabulated in Table IV. It is observed that with the P_E-criterion, there is a positive coding gain when using a biorthogonal code and the gain increases with the number of codewords. Again, with the P_b-criterion, there is always a coding loss and the loss increases with k.

For the (24, 12) extended Golay code,

$$G_E = \frac{2^{22}}{24 \cdot 12}\left(759\sqrt{8}P_8 + 2576\sqrt{12}P_{12}\right)^2 \approx 1.16 \approx 0.66 \text{ dB},$$

which is a gain over no coding. Also

$$G_b = \left(\frac{k}{n}\right)^2 G_E \approx 0.291 \approx -5.3 \text{ dB},$$

which is a loss. For the (15, 6) expurgated BCH code,

$$G_E = \frac{2^{10}}{15 \cdot 6}\left(30\sqrt{6}P_6 + 15\sqrt{8}P_8 + 18\sqrt{10}P_{10}\right)^2 \approx 1.35 \approx 1.3 \text{ dB},$$

and

$$G_b = \left(\frac{k}{n}\right)^2 G_E \approx 0.216 \approx -6.6 \text{ dB}.$$

It was shown in [1] that if hard quantization is used on an AWGN channel, using the bit-error probability criterion, any coding scheme results in a loss at low signal-to-noise ratios. Note that for all the codes discussed in the last section, based on the P_b-criterion, there is a loss with respect to no coding. We conjecture that this is true for binary codes in general on an unquantized AWGN channel with maximum-likelihood decoding. Since, at low signal-to-noise ratios, maximum-likelihood decoding is not the scheme that minimizes the bit-error probability, a stronger conjecture is that, based on the P_b-criterion, any coding scheme will result in a loss on an unquantized AWGN channel at sufficiently low signal-to-noise ratios.

VII. CONCLUDING REMARKS

In this paper, we have derived error probabilities and asymptotic coding gains of binary block codes used on an unquantized AWGN channel at very low signal-to-noise ratios. The results show that the performance depends heavily on the codes' global geometric structures through the important quantity P_i for each codeword. Since the computation of P_i involves degenerate multivariate normal random variables, for most cases it is very difficult to get closed-form expressions and we do not expect such exist. In order to have better insight to the performance of codes at low signal-to-noise ratios, we suggest further research in finding tight lower and upper bounds for P_i.

From the results of this paper, we know that the codewords which cover other codewords do not affect the low signal-to-noise ratio performance at all. We conjecture that for $\boldsymbol{x}_i, \boldsymbol{x}_j \in \boldsymbol{C}$, if $d_i < d_j$, then $P_i > P_j$. Now consider the quantity $\sum_{i=1}^{M-1}\sqrt{d_i}P_i$, which plays an important part in the expressions of error probabilities. It is observed that the term $\sqrt{d_i}P_i$ at the code's minimum distance is larger than the sum of remaining terms. We also conjecture that it is true in general.

One may wonder how the approximations obtained for the block-error probability and bit-error probability compare to the exact performances at low signal-to-noise ratios. An example is shown in Fig. 1, where the exact block-error probability and the approximation obtained in the paper for the biorthogonal code with $M = 16$ codewords on an unquantized AWGN channel are plotted as functions of the bit signal-to-noise ratio E_b/N_0. As seen from the figure, the approximation is very tight for E_b/N_0 up to −10 dB and not bad up to −5 dB.

APPENDIX A
PROOFS OF THEOREMS 1 AND 2

Theorem 1: Let $X_1, X_2, \cdots, X_{M-1}$ be $M - 1$ mean zero jointly normal random variables (possibly degenerate), with covariances σ_{ij} and with cumulative distribution function $F(x_1, x_2, \cdots, x_{M-1})$, with every pair X_i, X_j linearly independent (nondegenerate). Let $a_1, a_2, \cdots, a_{M-1}$ be nonnegative real numbers. Then, for $x > 0$,

$$F(a_1x, a_2x, \cdots, a_{M-1}x) = F(0, 0, \cdots, 0) + \frac{x}{\sqrt{2\pi}} \sum_{i=1}^{M-1} \frac{a_i}{\sqrt{\sigma_{ii}}} P_i + O(x^2), \quad (19)$$

where

$$P_i = \Pr\{X_j \le 0 \text{ for } j \ne i \mid X_i = 0\}$$

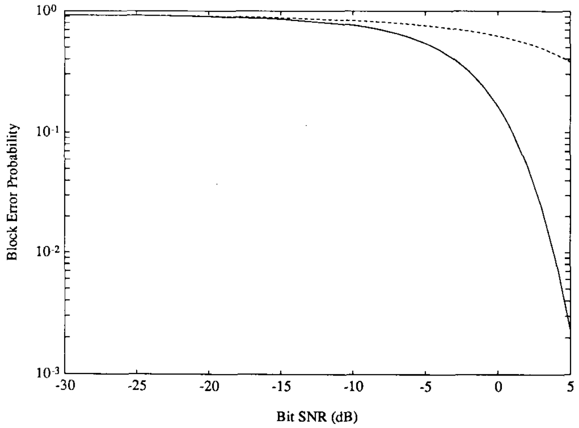

Fig. 1. Block-error probability of the biorthogonal code with $M = 16$ codewords on an unquantized AWGN channel. (Solid line: exact performance. Dashed line: approximation.)

$$\overset{\text{def}}{=} \lim_{h \to 0} \frac{\Pr\{X_j \le 0 \text{ for } j \ne i, 0 < X_i \le h\}}{\Pr\{0 < X_i \le h\}}.$$

Proof: We partition the event $\{X_1 \le a_1 x, \cdots, X_{M-1} \le a_{M-1} x\}$ into 2^{M-1} subsets, according to which of the X_i's are positive. If none of them are positive, we get the term $F(0, 0, \cdots, 0)$ in (19). If two or more of the X_i's are positive, say $X_i > 0$ and $X_j > 0$, then the corresponding probability is $\le \Pr\{0 < X_i \le a_i x, 0 < X_j \le a_j x\}$, which, since we assume that X_i and X_j are linearly independent, satisfies

$$\Pr\{0 < X_i \le a_i x, 0 < X_j \le a_j x\} = O(x^2),$$

so that the sum of the terms with at least two positive X_i's is $O(x^2)$. Finally we come to the terms with exactly one positive X_i. A typical term of this form is

$$\Pr\{X_j \le 0 \text{ for } j \ne i, 0 < X_i \le a_i x\}$$
$$= \Pr\{0 < X_i \le a_i x\} \cdot \Pr\{X_j \le 0 \text{ for } j \ne i \,|\, 0 < X_i \le a_i x\}. \tag{20}$$

As $x \to 0$, we have

$$\Pr\{0 < X_i \le a_i x\} = \frac{a_i x}{\sqrt{2\pi\sigma_{ii}}} + O(x^2),$$

and

$$\Pr\{X_j \le 0 \text{ for } j \ne i \,|\, 0 < X_i \le a_i x\} = P_i + O(x),$$

from Theorem 2 , which follows. Thus, the term (20) contributes $a_i P_i x / \sqrt{2\pi\sigma_{ii}} + O(x^2)$ to the formula (19). This completes the proof. □

Theorem 2: Define, for $h \ge 0$,

$$P_i(h) = \Pr\{X_j \le 0, \text{ for } j \ne i \,|\, 0 < X_i \le h\}.$$

Also, define the random variables Y_j, for $j \ne i$, by

$$Y_j = \sigma_{ii} X_j - \sigma_{ij} X_i.$$

Then,

$$\lim_{h \to 0} P_i(h) = P_i = \Pr\{Y_j \le 0, \text{ for } j \ne i\}. \tag{21}$$

Also,

$$|P_i(h) - P_i| = O(h). \tag{22}$$

Proof: If we can show that

$$\Pr\{Y_j \le -|\sigma_{ij}|h, \text{for } j \ne i\} \le P_i(h) \le \Pr\{Y_j \le |\sigma_{ij}|h, \text{for } j \ne i\}, \tag{23}$$

then (21) follows immediately. We first note that if $0 < X_i \le h$, for each $j \ne i$, we have the following set inclusions:

$$\{Y_j \le -|\sigma_{ij}|h\} \subseteq \{X_j \le 0\} \subseteq \{Y_j \le |\sigma_{ij}|h\}.$$

From this, it follows that

$$\Pr\{Y_j \le -|\sigma_{ij}|h, \text{ for } j \ne i \,|\, 0 < X_i \le h\} \le P_i(h)$$
$$\le \Pr\{Y_j \le |\sigma_{ij}|h, \text{ for } j \ne i \,|\, 0 < X_i \le h\}. \tag{24}$$

Now if $i \ne j$, X_i and Y_j are uncorrelated since

$$E(X_i Y_j) = \sigma_{ii} E(X_i X_j) - \sigma_{ij} E(X_i^2) = \sigma_{ii}\sigma_{ij} - \sigma_{ij}\sigma_{ii} = 0,$$

so that the conditions in (24) can be deleted, giving (23). To prove the estimate (22), note that the set difference $\{Y_j \le |\sigma_{ij}|h\} - \{Y_j \le -|\sigma_{ij}|h\}$ is contained in the union $\bigcup_{i=1}^{M-1}\{-|\sigma_{ij}|h \le Y_j \le |\sigma_{ij}|h\}$, so that the difference between the right and left sides of (23) is bounded by

$$\sum_{i=1}^{M-1} \Pr\{-|\sigma_{ij}|h \le Y_j \le |\sigma_{ij}|h\}. \tag{25}$$

And finally, each term in (25) is $O(h)$. This proves (22). □

Appendix B

Derivations of (9) and (10)

If the code C has no zero columns and no repeated columns, then it is easily shown that

$$\sum_{i=1}^{M\text{-}1} x_{ij}^2 = 2^{k-1}, \qquad \text{for } j = 1, 2, \cdots, n, \tag{26}$$

$$\sum_{i=1}^{M\text{-}1} x_{ij}\bar{x}_{ij} = 0, \qquad \text{for } j = 1, 2, \cdots, n, \tag{27}$$

$$\sum_{i=1}^{M\text{-}1} x_{ij}x_{il} = 2^{k-2}, \qquad \text{for } j, l = 1, 2, \cdots, n \text{ and } j \ne l, \tag{28}$$

$$\sum_{i=1}^{M\text{-}1} x_{ij}\bar{x}_{il} = 2^{k-2}, \qquad \text{for } j, l = 1, 2, \cdots, n \text{ and } j \ne l, \tag{29}$$

where x_{ij}, $j = 1, 2, \cdots, n$, are the components of $\boldsymbol{x}_i$ and $\bar{x}_{ij}$ is the complement of x_{ij}. Note that the symmetric assumption we made about each bit position of the codeword implies that there are no zero columns.

We have

$$\sum_{i=1}^{M\text{-}1} d_i^2 = \sum_{i=1}^{M-1} \left(\sum_{j=1}^{n} x_{ij}\right)\left(\sum_{m=1}^{n} x_{im}\right).$$

If $j = m$, then by (26)

$$\sum_{j=1}^{n} \sum_{i=1}^{M\text{-}1} x_{ij}^2 = n2^{k-1}.$$

On the other hand, if $j \neq m$, then by (28)

$$\sum_{j=1}^{n}\sum_{\substack{m=1\\ m\neq j}}^{n}\sum_{i=1}^{M-1} x_{ij}x_{im} = n(n-1)2^{k-2}.$$

Therefore,

$$\sum_{i=1}^{M-1} d_i^2 = n2^{k-1} + n(n-1)2^{k-2}$$
$$= n(n+1)2^{k-2}.$$

This ends the derivation of (9).

Similarly,

$$\sum_{i=1}^{M-1} d_i d(\boldsymbol{x}_i \oplus \boldsymbol{x}_l) = \sum_{i=1}^{M-1}\left(\sum_{j=1}^{n} x_{ij}\right)\left(\sum_{m=1}^{n}(x_{im}\oplus x_{lm})\right).$$

If $j = m$, then by (26) and (27)

$$\sum_{i=1}^{M-1} x_{ij}(x_{ij}\oplus x_{lj}) = \begin{cases} 2^{k-1}, & \text{if } x_{lj}=0,\\ 0, & \text{if } x_{lj}=1.\end{cases}$$

Since there are $(n-d_l)$ of x_{lj}'s such that $x_{lj}=0$,

$$\sum_{j=1}^{n}\sum_{i=1}^{M-1} x_{ij}(x_{ij}\oplus x_{lj}) = (n-d_l)2^{k-1}.$$

On the other hand, if $j \neq m$, then by (28) and (29)

$$\sum_{i=1}^{M-1} x_{ij}(x_{im}\oplus x_{lm}) = 2^{k-2}.$$

It follows that

$$\sum_{j=1}^{n}\sum_{\substack{m=1\\ m\neq j}}^{n}\sum_{i=1}^{M-1} x_{ij}(x_{im}\oplus x_{lm}) = n(n-1)2^{k-2}.$$

Finally,

$$\sum_{i=1}^{M-1} d_i d(\boldsymbol{x}_i \oplus \boldsymbol{x}_l) = (n-d_l)2^{k-1} + n(n-1)2^{k-2}$$
$$= n(n+1)2^{k-2} - d_l 2^{k-1},$$

which is the result of (10).

Appendix C

Proof of Theorem 5

The Farkas Alternative [10, p. 56]: Either the equation

$$\boldsymbol{Ax} = \boldsymbol{b} \quad \text{has a solution } \boldsymbol{x} \geq 0$$

or (exclusively)

$$\boldsymbol{y}^T\boldsymbol{A} \geq 0, \quad \boldsymbol{y}^T\boldsymbol{b} < 0 \quad \text{has a solution } \boldsymbol{y}.$$

Lemma 1: Either the equation

$$\boldsymbol{Ax} + \alpha\boldsymbol{d} = \boldsymbol{b} \quad \text{has a solution } \boldsymbol{x} \geq 0, \quad \alpha \in R \tag{30}$$

or (exclusively)

$$\boldsymbol{y}^T\boldsymbol{A} \leq 0, \quad \boldsymbol{y}^T\boldsymbol{d} = 0, \quad \boldsymbol{y}^T\boldsymbol{b} > 0 \quad \text{has a solution } \boldsymbol{y}. \tag{31}$$

Proof: Set the unconstrained $\alpha = u - v$ in (30) and require $u \geq 0$ and $v \geq 0$. Now this lemma follows from the Farkas Alternative. □

Theorem 5: Let the set $\mathscr{A} = \{\boldsymbol{x}\colon \ \boldsymbol{Ax} \leq 0, \ \boldsymbol{d}^T\boldsymbol{x} = 0\}$ be nonempty. The inequality $\boldsymbol{b}^T\boldsymbol{x} \leq 0$ holds for all $\boldsymbol{x} \in \mathscr{A}$, if and only if $\boldsymbol{b} = \boldsymbol{A}^T\boldsymbol{y} + \alpha\boldsymbol{d}$, for some $\boldsymbol{y} \geq 0$ and $\alpha \in R$.

Proof: The proof for the sufficient condition is straightforward. Now suppose the necessary condition is wrong, and assume that $\boldsymbol{b}$ is not in the form of $\boldsymbol{A}^T\boldsymbol{y} + \alpha\boldsymbol{d}$, $\boldsymbol{y} \geq 0$ and $\alpha \in R$. Then, the equation $\boldsymbol{A}^T\boldsymbol{y} + \alpha\boldsymbol{d} = \boldsymbol{b}$ does not have a solution $\boldsymbol{y} \geq 0$, $\alpha \in R$. Therefore, the case (30) of Lemma 1 is wrong, and we must have the alternative:

$$\boldsymbol{x}^T\boldsymbol{A}^T \leq 0, \quad \boldsymbol{x}^T\boldsymbol{d} = 0, \quad \boldsymbol{x}^T\boldsymbol{b} > 0 \quad \text{has a solution } \boldsymbol{x},$$

which contradicts the assumption that $\boldsymbol{b}^T\boldsymbol{x} \leq 0$ holds for all $\boldsymbol{x} \in \mathscr{A}$. □

References

[1] E. C. Posner, "Properties of error-correcting codes at low signal-to-noise ratios," *SIAM J. Appl. Math.*, vol. 15, pp. 775–798, July 1967.

[2] W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. 2, 2nd ed. New York: Wiley, 1971.

[3] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Amsterdam: North Holland, 1977.

[4] J. H. van Lint, *Introduction to Coding Theory*. New York: Springer-Verlag, 1982.

[5] W. W. Peterson and E. J. Weldon, Jr., *Error Correcting Codes*, 2nd ed. Cambridge, MA: MIT Press, 1972.

[6] J. S. Leon, "Computing automorphism groups of error-correcting codes," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 496–511, May 1982.

[7] J. J. Costa, "An algorithm for finding the automorphism group of a linear code," Ph.D. dissert., Univ. of Southern California, Los Angeles, 1985.

[8] C.-C. Lu, "The automorphism groups of binary primitive BCH codes," Ph.D. dissert., Univ. of Southern California, Los Angeles, 1987.

[9] J. Franklin, *Methods of Mathematical Economics*. New York: Springer-Verlag, 1980.

[10] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*. New York: McGraw-Hill, 1979.

[11] A. J. Viterbi, "On coded phase-coherent communications," *IRE Trans. Space Electron. Teleme.*, vol. SET-7, pp. 3–14, Mar. 1961.