An Upper Bound on the Error Probability of Linear Binary Block Codes in AWGN Interference

Shahram Yousefi[0]
Dept. of Elec. and Comp. Eng.
University of Waterloo
Waterloo, ON, Canada N2L 3G1
shahram@shannon.uwaterloo.ca

Amir K. Khandani
Dept. of Elec. and Comp. Eng.
University of Waterloo
Waterloo, ON, Canada N2L 3G1
khandani@shannon.uwaterloo.ca

Abstract — **In this article, a new upper bound for linear binary block codes based on Gallager's first bounding technique and an argument similar to random coding argument is proposed. This bound is very simple to calculate as it only requires the spectrum of the code. Also, it is particularly tight at low SNR's and for codes with large codebooks. Without much added complexity, the union bound is tightened.**

I. Introduction

The problem of performance evaluation of linear binary block codes with soft decision Maximum-Likelihood (ML) decoding has long been a central problem in coding theory and practice [2]–[12]. In most of the cases the derivation of a closed-form expression for, or even calculation of, the bit or word error probabilities are intractable, if not impossible. Thus, we usually resort to bounding techniques for the aforementioned probabilities.

The most commonly used upper bound on the error probability of a digital communication system is the *union bound.* Union bound is in fact an inequality from the class of *Bonferroni-type* [1] inequalities in probability theory. These are inequalities that are universally true regardless of the underlying probability space and for all choices of the basic events. For the calculation of the union bound on the error probability of a binary block code, one only needs to have the weight enumerating function (spectrum) of the code which results in much simplicity of calculation. The price to pay is, of course, accuracy. The union bound is quite accurate for high signal-to-noise ratios (SNR). For lower SNR's, however, union bound is a very poor upper bound.

For some applications such as concatenated coding schemes where the inner code is a binary block code, the low-SNR asymptotic coding gain of the code is needed for the performance evaluation of the overall scheme which explains the need to have tighter bounds at low SNR regions of the performance.

We present a new upper bound on the word error probability of binary block codes for all ranges of SNR. This is a very low-complexity bound based on an expansion of error probability similar to what Divsalar [2] refers to as "Gallager's first bounding technique". A random coding argument methodology is applied to the Gallager's method which results in an upper bound.

[0] This work was partly supported by Natural Sciences and Engineering Research Council of Canada (NSERC).

Consider a binary code[1] $C = \{c_0, c_1, ..., c_{2^k-1}\}$ with parameters (n,k,d), to be used along with Binary Phase Shift Keying (BPSK) modulation (antipodal signaling) on an Additive White Gaussian Noise (AWGN) channel. The resulting signal set will be $\mathcal{S} = \{s_0, s_1, ..., s_{2^k-1}\}$, where $s_i = \mathbf{m}(c_i) \in \mathbf{R}^n$. For $c_i = (c_{i1}, c_{i2}, ..., c_{i2^k-1})$, $\mathbf{m}(c_i) = (m(c_{i1}), m(c_{i2}), ..., m(c_{i2^k-1}))$. In the antipodal signaling case, $m(\alpha) = \sqrt{E_s}(2\alpha - 1)$, where E_s is the symbol energy.

As binary codes and binary modulation are "matched" in the sense of Loeliger [13], Euclidean distance which is the performance measure in the AWGN case will be proportional to Hamming distance[2]. In particular for BPSK, denoting the Euclidean distance between two signal points s_i and s_j by $\delta(s_i, s_j)$, we have:

$$\delta^2(s_i, s_j) = \|s_i - s_j\|^2 = 4E_s \cdot d(c_i, c_j) = 4RE_b \cdot d(c_i, c_j) \quad (1)$$

where $R = k/n$ is the binary code rate, $\|.\|$ is the usual Euclidean norm, and $d(,)$ is Hamming distance. Assuming AWGN interference, the output of the channel will be a vector $r = s_i + \mathbf{n}$, where $\mathbf{n}$ is an n-dimensional vector whose elements are independent zero-mean Gaussian random variables with a variance of $\sigma^2 = N_0/2$. Probability of word error for communicating one of 2^k messages in $\mathcal{S}$ through an AWGN channel will be:

$$P_w(E) = \sum_{i=0}^{2^k-1} P(E \mid s_i) P(s_i) \quad (2)$$

If the resulting "Geometrically-Uniform" (GU) signal set is equiprobable, the Maximum-Likelihood (ML) optimum decoding rule will actually reduce to minimum Euclidean distance decoding strategy and

$$P_w(E) = P(E \mid s_i) \quad (3)$$

where s_i can be any signal point. We assume that s_0, signal corresponding to the all-zero codeword, has been transmitted.

The difficulty in calculating $P(E \mid s_i)$ is due to the complexity of the *decision* or *Voronoi* regions of the signal points; that of s_0 in particular for this matter. Voronoi region of any signal point in an n-dimensional signal constellation is the set of points or vectors in $\mathbf{R}^n$ that are closer to that point than any other point in the constellation. i.e.,

$$\mathcal{V}_i = \{x \in \mathbf{R}^n : \delta(x, s_i) \leq \delta(x, s), \forall s \in \mathcal{S}\} \quad (4)$$

[1] In this correspondence by a binary code we mean a linear binary block code.

[2] This proportionality does not hold in general for other signal sets.

$\mathcal{V}_i$ in general is the volume bounded by $M = 2^k - 1$ inequalities in 4, each of which specifying a half-space located on one side of a hyper-plane defined by the inequalities of 4. The resulting region is a convex polytope in $\mathbf{R}^n$ and

$$P(E \mid s_i) = P(r \in \mathcal{V}_i^c \mid s_i) = 1 - P(r \in \mathcal{V}_i \mid s_i). \quad (5)$$

In order to find the exact probability of error, one needs to integrate an n-dimensional Gaussian probability density function over the polytope $\mathcal{V}_i$. This can be a challenging problem if the angles between the facets of the polytope are not right angles. For the 2-dimensional case Craig [3] cleverly rephrased the resulting double integral and came up with an exact expression for the error probability in terms of a sum of a few single integrals of elementary functions with finite ranges.

The amount of work done on lower and upper bounding techniques and approximations to the the word and bit error probabilities in specific SNR ranges is overwhelming.

Union bound is a Bonferroni-type inequality capping the probability of union of any arbitrary set of events like $A_1, A_2, ..., A_M$, as $P\left(\bigcup_{j=1}^{M} A_j\right) \leq \sum_{j=1}^{M} P(A_j)$.

The word error probability, $P_w(E) = P\left(\bigcup_{j=1}^{M} E_j\right)$, where $E_j = \{\|r - s_j\| \leq \|r - s_0\| | s_0\}$, for $j = 1, 2, ..., M$; is then bounded by,

$$P_w(E) \leq \sum_{j=1}^{M} E_j = \sum_{j=1}^{M} Q\left(\frac{\delta_{j0}}{2\sigma}\right)$$

where $\delta_{j0} = \delta(s_j, s_0)$, and $Q(.)$ is the well known Q-function:

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-u^2/2} du.$$

The above union (or additive) bound is asymptotically tight. The Euclidean weight enumerator of the signal set comes directly from the Hamming weight enumerator of the underlying binary code as:

$$ewe(w) = \sum_{j=1}^{\eta} A_j w^{\delta_{j0}} \quad (6)$$

in which $\eta \leq M$ is the number of Euclidean layers or shells (simply the number of distinct Euclidean distances from any reference point, say s_0) of the signal set and A_j is the number of signal points in the j-th shell.

Then,

$$P_w(E) \leq \sum_{j=1}^{\eta} A_j Q\left(\frac{\delta_{j0}}{2\sigma}\right). \quad (7)$$

For low probability of error (high SNR) region, sometimes only the first term of the spectrum, i.e., the term corresponding to the minimum distance of the code, suffices to provide a good approximation to $P_w(E)$. For higher error probabilities, this bound becomes too poor to be acceptable.

One of the first works devoted to the performance of binary codes at low signal-to-noise ratios is that of Posner [4] which mainly revolves around quantized channel, i.e., with hard decision. A belated continuation to the work of Posner for the un-quantized channel output (soft decision) is that of Chao et. al [5]. In [5], a power series expansion of the probability of correct decision around zero SNR is used to climax at a relatively accurate, albeit complex, approximation to the word error probability. The complexity of their result is due to the fact that their expression for the error probability is a function of a parameter which depends on the "global" geometrical properties of the code.

One other approach to bounding is to approximate the polytope decision region by a simpler region; usually spherical or conical. These two "work" because the probability density function, $P_{\mathbf{n}}(x)$, of a Gaussian vector $\mathbf{n}$ is monotonically decreasing with $\|x\|$ and is independent of the angle orientation of the x. A spherical approximation leads to the *minimum distance bound* [6] which is only good for small values of n (dimensionality). One important improvement to the minimum distance and union bound is that of Hughes [7]. In Hughes' work a conical approximation is used. The complement of the decision region for the reference point is decomposed into the union of a set of disjoint polyhedral cones. Approximating these polyhedral cones with circular cones with the same solid angle results in an upper bound.

Hughes work launched a number of similar works with applications from linear binary block codes to coded modulation and concatenated codes both in AWGN and fading environments. Many other bounds, as noted by Divsalar [2], "essentially use a general bounding technique developed by Gallager" [8]. Namely, given a transmitted signal,

$$\begin{aligned} P_w(E) &= P\{\text{word error}, r \in \Re\} + P\{\text{word error}, r \notin \Re\} \\ &= P\{\text{E}, r \in \Re\} + P\{\text{E} \mid r \notin \Re\} \cdot P\{r \notin \Re\} \\ &\leq P\{\text{E}, r \in \Re\} + P\{r \notin \Re\} \end{aligned} \quad (8)$$

where r is the received vector signal and $\Re$ is an appropriate region around the transmitted signal point. Divsalar [2] refers to this as "Gallager's first bounding technique". In original Gallager's work $\Re$ is a complicated region in $\mathbf{R}^n$. Simpler bounds are developed where this region is approximated by polyhedral or circular cones or even hyper-spheres.

Herzberg and Poltyrev [9] use Gallager's first bounding technique in 8 to derive one of the tightest upper bounds. $\Re$ is chosen to be a hyper-sphere with radius r and then the bound is tightened over r. They apply their method to Block-Coded Modulation (BCM) schemes communicated over AWGN channel.

The Tangential Sphere Bound (TSB) proposed for binary codes in [10] and for MPSK BCM schemes in [11] also uses Gallager's first bounding technique where $\Re$ is a conical region.

Another line of attack to the bounding problem of error probability is use of Bonferroni-type inequalities. This method accounts for a number of tight upper as well as lower bounds on the error probability.

A lot of the proposed bounds are either too complex or too loose at low SNR's. Their complexity is due to-among many others-dependence on global properties of the underlying code or complex optimization over one or more parameters or even a complex equation to be solved.

II. Our Random-Hyper-plane (RHP) Upper Bound

Our upper bound is primarily based on the Gallager's first bounding technique given in 8. The choice of region $\Re$ is of utmost significance in this bounding method. Different choices of this region have resulted in various different tight bounds in different ranges of signal-to-noise ratio. In the proposed bound, we define $\Re^c$ as the union of ℓ regions such that $\Re^c = \bigcup_{j=1}^{\ell} H_j$, i.e., $\Re = \bigcap_{j=1}^{\ell} H_j^c$; where regions H_j are each a half n-dimensional space defined as the set of all the points in the space which are closer (in Euclidean sense) to a random point v_j than s_0. i.e.,

$$H_j = \{r \in R^n : \|r - v_j\| < \|r - s_0\|\} \tag{9}$$

The perpendicular bisector of the line joining the points v_j and s_0 is an $(n-1)$-dimensional hyper-plane in the n-space and therefore $P(H_j)$ is simply the probability that the received vector r has passed this hyper-plane.

We randomly select vectors v_j whose n coordinates are binary distributed over $\{-1, +1\}$, i.e., $P(v_{ij} = +1) = P(v_{ij} = -1) = 1/2$. That is, all the signal points as well as the random vectors are on an n-sphere with Euclidean radius n.

Then,

$$P\{r \notin \Re\} = P\{\bigcup_{j=1}^{\ell} H_j\} \leq \sum_{j=1}^{\ell} Q\left(\frac{\sqrt{d(s_0, v_j)}}{\sigma}\right) \tag{10}$$

where $d(s_0, v_j)$ is the hamming distance between s_0 and v_j, that is the number of coordinates that s_0 and v_j are different.

For the sake of brevity of the bound, we impose one other constraint on the random vectors v_j: $\forall j, d(s_0, v_j) = w_r$, i.e., all the random vectors are at the same Euclidean distance from the transmitted signal. Also, as $E = \bigcup_{i=1}^{M} E_i$,

$$P\{\mathrm{E}, r \in \Re\} = P\left\{\bigcup_{i=1}^{M} E_i, \bigcap_{j=1}^{\ell} H_j^c\right\} = P\left\{\bigcup_{i=1}^{M}\left(E_i, \bigcap_{j=1}^{\ell} H_j^c\right)\right\} \tag{11}$$

and,

$$P\left\{\bigcup_{i=1}^{M}\left(E_i, \bigcap_{j=1}^{\ell} H_j^c\right)\right\} \leq \sum_{i=1}^{M} P\left\{E_i, \bigcap_{j=1}^{\ell} H_j^c\right\} \leq \sum_{i=1}^{M} P\left\{E_i, H_{j_{opt,i}}^c\right\}$$

where for each i , $j_{opt,i} \in \{1, 2, ..., \ell\}$ is the index of the hyper-plane for which $P\left\{E_i, H_j^c\right\}$ is minimum. The joint probability $P\left\{E_i, H_j^c\right\}$ is a 2-dimensional Gaussian probability and is shown to be (see appendix):

$$\begin{aligned} P\left\{E_i, H_j^c\right\} &= \frac{1}{2\pi}\int_{-\pi/2}^{\theta_{max}} \exp\left(-\frac{-a^2}{2\sigma^2\cos^2(\theta)}\right) d\theta \\ &-\frac{1}{2\pi}\int_{\phi-\pi/2}^{\theta_{max}} \exp\left(-\frac{-b^2}{2\sigma^2\cos^2(\phi-\theta)}\right) d\theta \end{aligned} \tag{12}$$

where $a = \delta(s_0, s_i)/2$, $b = \delta(s_0, v_j)/2$, $\phi = \arccos(\rho)$, and $\theta_{max} = \arctan\left(\frac{b - a\cos\phi}{a \sin\phi}\right)$. The correlation coefficient ρ is shown to be equal to:

$$0 \leq \rho = \frac{d(c_0, c_i) + d(c_0, m^{-1}(v_j)) - d(s_i, m^{-1}(v_j))}{2\sqrt{d(c_0, c_i) \cdot d(c_0, m^{-1}(v_j))}} \leq 1 \tag{13}$$

Representing $d(c_0, c_i) = d_i$ and $d(c_0, m^{-1}(v_j)) = w_r$, $\rho = (d_i + w_r - d(s_i, m^{-1}(v_j)))/(2\sqrt{d_i \cdot w_r})$ will be a positive value (as Hamming distance satisfies the triangular inequality) whose larger values would result is smaller values of $P\left\{E_i, H_j^c\right\}$ as desired because for fixed a and b, the 2-dimensional Gaussian integral in 12 is a monotonically decreasing function of ρ. As vectors v_j are random, so are the correlation coefficients. The distribution of ρ for $k_{min} \leq k \leq k_{max}$ is,

$$P_\rho\left(\rho = \frac{k}{d_i \cdot w_r}\right) = \frac{\binom{d_i}{k}\binom{n-d_i}{w_r - k}}{\binom{n}{w_r}} \tag{14}$$

where $k_{min} = \max(0, d_i + w_r - n)$ and $k_{max} = \min(d_i, w_r)$. Denoting $P_x[k] = P_x\left(\rho_{max} = \frac{k}{d_i \cdot w_r}\right)$, the best ρ (i.e., ρ_{max}) out of the ℓ possibilities will have the following distribution:

$$P_{\rho_{max}}[k] = \left(\sum_{q=k_{min}}^{k} P_\rho[q]\right)^\ell - \left(\sum_{q=k_{min}}^{k-1} P_\rho[q]\right)^\ell \tag{15}$$

Therefore, the overall random hyper-plane upper bound will be:

$$P_w(E) \leq \min_{w_r, \ell}\left(\ell \cdot Q\left(\frac{\sqrt{w_r}}{\sigma}\right) + \sum_{i=1}^{M} \overline{P\left\{E_i, H_{j_{opt,i}}^c\right\}}^{\rho_{max}}\right) \tag{16}$$

where the bar is an averaging on ρ_{max}. This optimization is done over $1 \leq w_r \leq n$ and ℓ. The tradeoff regarding ℓ is obvious from the above bound. Higher values of ℓ will result in an increase in the first term in 16 but as they increase the probability of higher ρ_{max}'s they would subside the second term in 16. One significant advantage of the proposed bound is that the optimization must be done in the specific SNR of interest which in return will result in a tight bound for that point of interest.

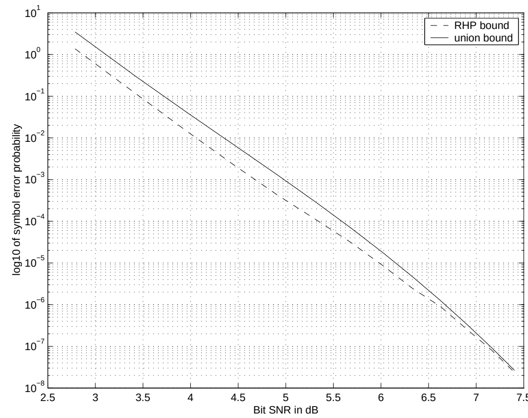

Figure 1: Symbol error probability for BCH(63,51).

III. Numerical Results and Discussion

The random hyper-plane bound proposed is specifically tight for higher dimensional codes. For smaller codebooks the first term of the bound in 16 becomes so significant that it completely cancels out the advantage gained in the second term.

The proposed bound applied to the BCH$(63, 51)$ code along with the union bound are shown in Fig. 1. Optimizing values at a symbol error probability of 10^{-2} are $w_r = 19$ and $\ell = 3 \times 10^8$. Different values of w_r and ℓ result is slight improvement at other SNR's with respect to the shown RHP bound in Fig. 1.

IV. Appendix: Proof of 12

$P\left\{E_i, H_j^c\right\}$ is a 2-dimensional Gaussian probability as in:

$$P(E_i \cap H_j^c) = P(r \in R_I) = \int_{\delta_{i0}/2}^{\infty} \int_{-\infty}^{\delta_{j0}/2} f(z_i, z_j; \sigma^2, \rho_{ij}) dz_i dz_j \tag{17}$$

where R_I is the shaded infinite sector in Fig. 2 and,

$$f(z_i, z_j; \sigma^2, \rho_{ij}) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho_{ij}^2}} . \exp\left(-\frac{z_i^2 + z_j^2 - 2\rho_{ij} z_i z_j}{2(1-\rho_{ij}^2)\sigma^2}\right)$$

and $\delta_{i0} = \delta(s_i, s_0)$, $\delta_{j0} = \delta(v_j, s_0)$, and the Gaussian variables z_i and z_j are the components of noise in the $(s_i - s_0)$ and $(v_j - s_0)$ directions, respectively(see Figure 2). i.e.,

$$z_i = < \mathbf{n}, (s_i - s_0) > \quad , \quad z_j = < \mathbf{n}, (v_j - s_0) > \tag{18}$$

where $\mathbf{n}$ is the Gaussian noise vector and $< . >$ is the inner product operator. The correlation coefficient between z_i and z_j, ρ_{ij} is

$$\rho_{ij} = \frac{< (s_i - s_0), (v_j - s_0) >}{\|s_i - s_0\| . \|v_j - s_0\|}. \tag{19}$$

It is trivial to see that

$$\rho_{ij} = \frac{d_{i0} + d_{j0} - d_{ij}}{2\sqrt{d_{i0}.d_{j0}}} \tag{20}$$

where d_{ij} is, as defined before, the Hamming distance between the underlying codewords of the binary code or equivalently the "time diversity" between the corresponding signals s_i and s_j or between s_i and v_j in the case of a hyper-plane, i.e., $d_{i0} = d(c_i, c_0)$, $d_{j0} = d(m^{-1}(v_j), c_0)$, and $d_{ij} = d(c_i, m^{-1}(v_j))$.

Using a method similar to that of [3] and [12], we can change the double integral with infinite limits to a single integral with finite limits. We start with writing the 2-dimensional Gaussian integral in 17 in terms of the polar form of the 2-dimensional Gaussian function, i.e.,

$$P(r \in R_I) = \underbrace{\int\int}_{\mathrm{R_I}} f_p(\xi, \theta; \sigma^2, \rho) d\xi d\theta \tag{21}$$

where

$$f_p(\xi, \theta; \sigma^2, \rho) = \frac{\xi}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp\left[-\frac{\xi^2}{2\sigma^2(1-\rho^2)}(1 - \rho\sin 2\theta)\right]$$

and therefore it is easy to see from the geometry of the problem that

$$\begin{aligned} P(r \in R_I) &= \int_{-\pi/2}^{\theta_c} \int_{\xi_1}^{+\infty} f_p(\xi, \theta; \sigma^2, \rho) d\xi d\theta \\ &+ \int_{\theta_c}^{\theta_{max}} \int_{\xi_1}^{\xi_2} f_p(\xi, \theta; \sigma^2, \rho) d\xi d\theta \end{aligned} \tag{22}$$

where from $\triangle ACs_0$ and $\triangle BDs_0$ we find $\xi_1 = \bar{s_0 A} = \frac{a}{\cos\theta}$ and $\xi_2 = \bar{s_0 B} = \frac{b}{\cos(\phi - \theta)}$, respectively. The maximum possible angle can be obtained from the equation $\xi_1 = \xi_2$ as $\theta_{max} = \tan^{-1}\left(\frac{b - a\cos\phi}{a\sin\phi}\right)$. With these in hand, 22 can be simplified to 12.

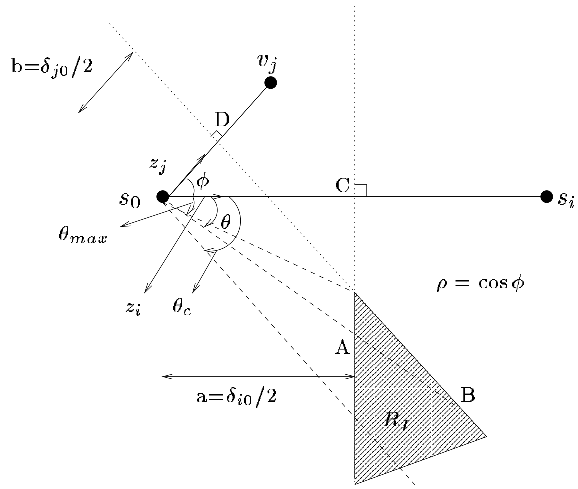

Figure 2: Geometry for $P(E_i \cap H_j^c)$.

References

[1] J. Galambos, and I. Simonelli, *Bonferroni-type Inequalities with Applications*, Springer 1996.

[2] D. Divsalar, "A simple tight bound on error probability of block codes with application to Turbo codes," *TMO Progress Report 42-139*, NASA, JPL, Pasadena, CA, USA, 1999.

[3] J. W. Craig, "A new, simple and exact result for calculating the probability of error for two-dimensional signal constellations," in *Proc. IEEE Military Commun. Conf.*, pp. 571-575, McLean, VA, USA, Oct. 1991.

[4] 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.

[5] C.-C. Chao, R. J. McEliece, L. Swanson, and E. R. Rodemich, "Performance of binary block codes at low signal-to-noise ratios," *IEEE Trans. Inform. Theory*, vol. IT-38, no. 6, pp. 1677-1687, Nov. 1992.

[6] C. E. Shannon, "Probability of error for optimal codes in Gaussian channel," *Bell Syst. Tech. J.*, vol. 38, pp. 611-656; and *Math Rev.*, vol. 21, p. 1920, 1959.

[7] B. Hughes, "On the error probability of signals in additive white Gaussian noise," *IEEE Trans. Inform. Theory*, vol. IT-37, no. 1, pp. 151-155, Jan. 1991.

[8] R. G. Gallager, " A simple derivation of the coding theorem and some applications," *IEEE Trans. Inform. Theory*, pp. 3-18, Jan. 1965.

[9] H. Herzberg, and G. Poltyrev, "Techniques of bounding the probability of decoding error for block-coded modulation structures," *IEEE Trans. Inform. Theory*, vol. IT-40, no. 3, pp. 903-911, May. 1994.

[10] G. Poltyrev, "Bounds on the decoding error probability of binary linear codes via their spectra," *IEEE Trans. Inform. Theory*, vol. IT-40, no. 4, pp. 1284-1292, July. 1994.

[11] H. Herzberg, and G. Poltyrev, "The error probability of M-ary PSK block-coded modulation schemes," *IEEE Trans. Commun.*, vol. COM-44, no. 4, pp. 427-433, Apr. 1996.

[12] M. K. Simon, and D. Divsalar, "Some new twists to problems involving the Gaussian probability integral," *IEEE Trans. Commun.*, vol. COM-46, no. 2, pp. 200-210, Feb. 1998.

[13] H.-A. Loeliger, "Signal sets matched to groups," *IEEE Trans. Inform. Theory*, vol. IT-37, no. 6, pp. 1675-1682, Nov. 1991.