Generalized Tangential Sphere Bound on the ML Decoding Error Probability of Linear Binary Block Codes in AWGN Interference

Shahram Yousefi and Amir K. Khandani

University of Waterloo

Coding and Signal Transmission (CST) Laboratory

Department of Electrical & Computer Engineering

Waterloo, Ontario, Canada, N2L 3G1

Technical Report UW-E&CE#2002-2

Feb. 18, 2002

Generalized Tangential Sphere Bound on the ML Decoding Error Probability of Linear Binary Block Codes in AWGN Interference

Shahram Yousefi and Amir K. Khandani

Abstract

The error probability of Maximum-Likelihood (ML) soft-decision decoded binary block codes rarely accepts nice closed forms. In addition, for long codes ML decoding becomes prohibitively complex. Nevertheless, bounds on the performance of ML decoded systems provide insight into the effect of system parameters on the overall system performance as well as a measure of goodness of the sub-optimum decoding methods used in practice. Using the so-called Gallager's first bounding technique (involving a so-called Gallager region) and within the framework of Tangential Sphere Bound (TSB) of Poltyrev, we develop a general bound referred to as the Generalized Tangential Sphere Bound (GTSB). The Gallager region is chosen to be a general Hyper-Surface of Revolution (HSR) which is optimized to tighten the bound. The search for the optimal Gallager region is a classical problem dating back to Gallager's thesis in the early 1960's. For the random coding case, Gallager provided the optimal solution in a closed form while for the non-random case the problem has been an active area of research in information theory for many years. We prove that for a sphere code the optimum HSR within the proposed GTSB is a hyper-cone. This will climax to the TSB of Poltyrev, one of the tightest bounds ever developed for binary block codes, and therefore terminates the search for a better Gallager region in the groundwork of the GTSB.

Keywords

Block codes, maximum-likelihood decoding, upper bounds, Gallager bounds, union bound.

This work was supported in part by Canadian Institute for Telecommunications Research (CITR) and in part by Natural Sciences and Engineering Research Council of Canada (NSERC). This work has been also submitted to the 21st biennial symposium on communications, Queen's University, Kingston, ON, Canada.

The authors are with Coding & Signal Transmission (CST) laboratory at the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada N2L 3G1 (email: shahram,khandani@shannon.uwaterloo.ca).

I. Introduction

The problem of performance evaluation of linear binary block codes with soft decision Maximum-Likelihood (ML) decoding in Additive White Gaussian Noise (AWGN) interference has long been a central problem in coding theory and practice. In most of the cases, the derivation of a closed-form expression for the bit or word error probabilities is intractable. Thus, one usually resorts 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. There are various Bonferroni-type upper as well as lower bounds exploited in communication theory such as KAT bound by Kuai et al. [2] (also see references in [2]). 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 union bound is quite accurate for high SNR's while for other SNR's, it 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 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.

Also, for longer binary block codes ML decoding becomes prohibitively complex. Within this context, tighter upper bounds on the ML decoding performance of binary block codes used in conjunction with Binary Phase Shift Keying (BPSK) modulation will provide means of assessing the performance of these codes.

The recent overwhelming attention given to the bounding techniques for performance evaluation of codes is mainly due to the introduction of some near-Shannon-limit performing schemes. Turbo codes, invented by Berrou et al. [3] in 1993, Repeat-Accumulate (RA) codes of Divsalar et al. [4], and Low Density Parity Check (LDPC) codes of Gallager [5], resurrected by MacKay et al. [6] in 1996 are the best examples. In addition to simulations, the aforementioned schemes can be analyzed using the union bound which is a very loose measure of performance for rates above the cutoff rate of the channel [7]. Therefore, there is an increasing demand for tighter bounds on the ML decoding of such codes above the cutoff rate.

The complexity of the calculation of error probability for ML-decoded BPSK-modulated binary codes is mainly due to the complexity of the so-called Voronoi or decision regions [8]. In order to find the probability of correct decision, one needs to integrate a multidimensional Gaussian distribution over the Voronoi region of the transmitted codeword. One of the first works devoted to the performance of binary codes at low signal-to-noise ratios is that of Posner [9] 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 [10]. In this work, a power series expansion of the probability of correct decision around zero SNR is used to compute 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 important improvement to the union bound is that of Hughes [11]. Hughes represented the complement of the Voronoi region (all Voronoi regions are congruent to each other for Slepian codes[1] [12, 13]) as the union of a set of truncated polyhedral cones and then cleverly upper-bounded the error probability by replacing those truncated cones by truncated right circular cones with the same solid angle for which the corresponding probabilities are larger but can easily be evaluated. Since the codewords of a binary code[2] used with binary modulation are not spread uniformly on the Euclidean sphere, Hughes' bound cannot be asymptotically tight. 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 [15–17].

II. Bounds based on Gallager's first bounding technique

Many other bounds, as noted by Divsalar [18], "essentially use a general bounding technique developed by Gallager [5]". In this method, Gallager bisects the error probability

[1] A Slepian signal set is a Geometrically-Uniform (GU) [14] and equi-energy (sphere) signal set.

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

to joint probability of error and noise residing in a region $\Re$ (referred to as the *Gallager region*) plus joint probability of error and noise residing in the complement of $\Re$; where $\Re$ is a volume around the transmitted codeword. Divsalar [18] refers to this as "Gallager's First Bounding Technique" (GFBT). In original Gallager's work $\Re$ is a complicated region in $\mathbf{R^n}$.

For instance, the well-known Tangential Bound (TB) of Berlekamp [19] uses Gallager's first bounding technique combined with union bound to provide a significantly tightened bound than the conventional union bound at low SNR's. This is achieved by separating the radial and tangential components of the Gaussian noise with a half-space as the underlying Gallager region.

Herzberg and Poltyrev [15] use GFBT 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. This is referred to as the Sphere Bound (SB) of Herzberg et al. They also apply their method to Block-Coded Modulation (BCM) schemes communicated over AWGN channel. BCM schemes involving MPSK (M-ary Phase Shift Keying) constellations are analogous to binary codes along with BPSK modulation as both are sphere signal sets, i.e., all the signal points reside on the surface of a hyper-sphere and therefore have the same energy.

The Tangential Sphere Bound (TSB) proposed for binary codes by Poltyrev [16] and for MPSK BCM schemes by Herzberg et al. [17] also uses GFBT where $\Re$ is a conical region. It is proven in [17] that the Berlekamp's TB is not tighter than TSB and de facto TSB is one of the tightest bounds to-date.

The tightening of the upper bounds on the ML decoding error probability of binary block codes within the format of the GFBT has been an evolutionary process: an evolution of the Gallager region from a half-space in the TB (Fig. 1-a) to a sphere in the SB (Fig. 1-b) and eventually to a cone in the TSB (Fig. 1-c). As it can be seen from Fig. 1, the common point between all of these regions is their azimuthal symmetry along the radial axis, namely z_1 here. For sphere codes, this is essentially the axis joining the transmitted signal on the surface of a sphere to the center of the sphere at the origin. *As a result, the cross sections of the Gallager region along the symmetry axis are spheres.* The difference between these regions, however, stems from the fact that the aforementioned cross sections

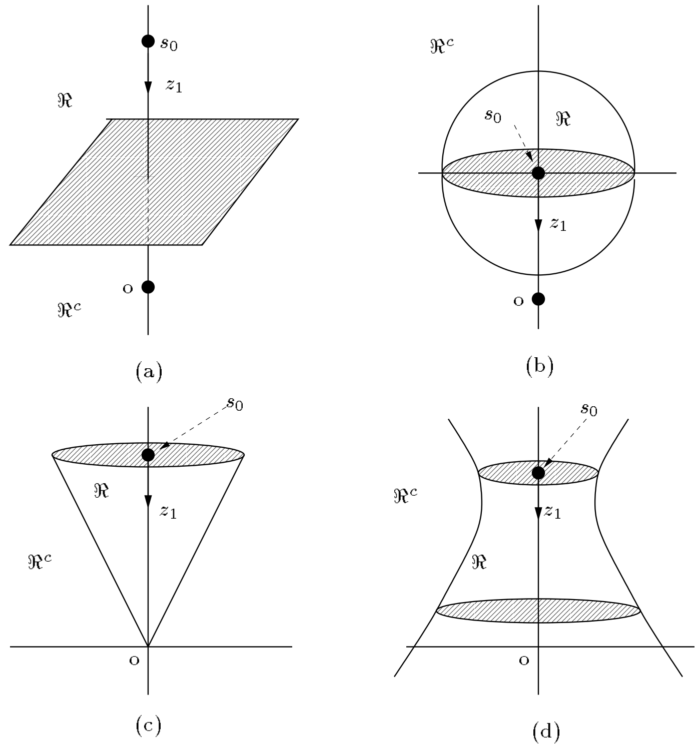

Fig. 1. Gallager regions: a) a half-space in the TB, b) a hyper-sphere in the SB, c) a hyper-cone in the TSB, and d) a hyper surface of revolution in the GTSB. z_1 is the azimuthal symmetry axis.

radii are different functions of z_1; the best of which being a linear function of z_1 in the TSB. However, the question still remained unanswered as to what boundary or region would result in the tightest bound in this formulation.

In this article, we extend the TSB to the so-called Generalized Tangential Sphere Bound (GTSB) by generalizing the Gallager region to a generic one encompassing all of the above cases. This will be a so-called Hyper Surface of Revolution (HSR) [20, 21] shown in Fig. 1-d and explained in the following section.

Using variational calculus, we obtain the optimal Gallager region within the resulting framework. This is shown to be a a right circular hyper-cone which coincides with the TSB of Poltyrev. This has long been an important open problem going back to Gallager's thesis since early 1960's where within his first bounding technique, GFBT, he introduced a function of observation space (denoted by $f(y)$ in the original work) to be optimized to tighten the bound [5, 7]. All other versions of the upper bounds based on GFBT (even those with one or more optimization parameters) are only asymptotically tight for random codes (as they achieve the capacity limits as $n \to \infty$). For nonrandom codes, as the underlying Gallager regions (which are optimized for random codes) are not optimum, the proposed bounds are not tight [18]. Using variational calculus, Gallager found the optimum $f(y)$ which was not in a closed form but reduced to a closed form for random codes. Divsalar [18] shows that the optimization of the Gallager region within the GFBT is equivalent to that of $f(y)$.

Albeit classically deemed an important problem, there has not been any mathematical proof for the optimality of the cone in the framework of the TSB. This proof is doubly important thanks to the wide spread applications of the TSB in various schemes. The convenience of relying solely on the code spectrum besides its extra tightness for lower rate codes, has made the TSB a good candidate for longer codes such as Turbo codes and LDPC codes.

Sason and Shamai [22] elaborated on TSB and applied it to parallel and serial concatenated Turbo codes using their ensemble spectrum and also extended the bound from word error probability to bit error probability. Their contribution to the TSB of Poltyrev is three-fold. First, they reestablished the validity of the bound by showing that for all instances of practical interest the probability of the lower half cone or lower nappe (i.e., when the radial component of the noise is smaller than negative of the root square of the signal energies) is negligible compared to the upper nappe and the overall error probability. Second, they provided rigorous proof for the existence and uniqueness of a solution for the optimization equation involved. And third, they prove that it is advantageous to apply the TSB to the whole codebook as opposed to the partitioned codebook as any partitioning of the signal set will yield looser results.

TSB has also been applied to LDPC codes [23] as well as to block codes communicated over interleaved fading channels [24].

III. Preliminaries

Consider a binary code $C = \{c_0, c_1, ..., c_{2^k-1}\}$ with parameters (n, k, d_{min}), to be used along with BPSK modulation (antipodal signaling) on an 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 $c_i = (c_{i1}, c_{i2}, ..., c_{i2^k-1})$,

$$\mathbf{m}(c_i) = (m(c_{i1}), m(c_{i2}), ..., m(c_{i2^k-1}))$$

where $m(\alpha) = \sqrt{E_s}(2\alpha - 1)$, $\alpha \in \{0, 1\}$, and E_s is the symbol energy[3]. The resulting signal set is a sphere and Slepian signal set.

As binary codes and binary modulation are *matched* in the sense of Loeliger [25], Euclidean distance which is the performance measure in the AWGN case will be proportional to Hamming distance[4]. In particular for BPSK, denoting the Euclidean distance between two signal points s_i and s_j by $\delta(s_i, s_j)$ or simply δ_{ij}, we have:

$$\delta_{ij}^2 = \delta^2(s_i, s_j) = \|s_i - s_j\|^2 = 4E_s d(c_i, c_j) = 4RE_b d(c_i, c_j) \tag{1}$$

where $R = k/n$ is the binary code rate, $\|.\|$ is the usual Euclidean norm, E_b is the information bit energy, 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 σ^2. Probability of word error for communicating one of 2^k messages in $\mathcal{S}$ through an AWGN channel will be:

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

If the resulting *Geometrically-Uniform* (GU) signal set [14] is equi-probable, the ML optimum decoding rule will actually reduce to minimum Euclidean distance decoding strategy and

$$P_w(E) = P(E \mid s_i) \tag{3}$$

where s_i can be any signal point. We assume that s_0, signal corresponding to the all-zero codeword, c_0, has been transmitted.

The difficulty in calculating $P(E \mid s_i)$ is due to the complexity of the *decision* or *Voronoi* regions [8] of the signal points which are convex polytopes in $\mathbf{R}^n$ [26].

[3]Without loss of generality, E_s is chosen to be unity.

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

IV. Generalized Tangential Sphere Bound Using A Hyper-Surface of Revolution

GTSB is primarily based on the Gallager's first bounding technique. Given a transmitted signal, the word error probability can be decomposed as in

$$\begin{aligned} P_w(E) &= P\{E, r \in \Re\} + P\{E, r \notin \Re\} \\ &= P\{E, r \in \Re\} + P\{E \mid r \notin \Re\} \cdot P\{r \notin \Re\} \\ &\leq P\{E, r \in \Re\} + P\{r \notin \Re\} \end{aligned} \tag{4}$$

where r is the received signal vector and $\Re$, referred to as the Gallager region, is an appropriate region around the transmitted signal point. 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. Examples of the Gallager region which have resulted in the tightest bounds include spheres [15] and right circular cones [16]. Motivated by the sensitivity of the bounds on the choice of Gallager region, we seek to find an optimum volume within the discussed playground while keeping the bound analytically tractable. In general, to have a tight bound for all ranges of signal-to-noise ratio, one would like to choose a region $\Re$ which is as close as possible-in geometrical sense-to the Voronoi regions. Our general bound is primarily for a sphere code of which all of the Voronoi regions are polyhedral cones having a single vertex at the origin of the n-space and extending infinitely in some direction [13, 26].

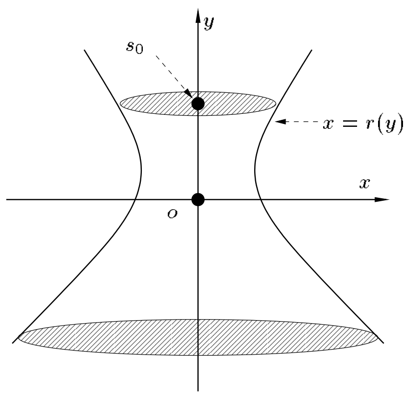

Fig. 2. Geometry of a surface of revolution: $x = r(y)$ rotates about the y axis to produce a surface with azimuthal symmetry along y.

At this point, as GTSB is developed geometrically, we start with an introduction to the geometry (analytic) required for the bound to transpire. GTSB, similar to TSB, is structured based upon the premise of multiple levels of separation of noise components from the rest of the noise vector, the first of which being the radial component of the noise. This is the projection of the noise vector along the $\overrightarrow{s_0 o}$ (see Fig. 2). The simplicity of the TB, SB, and TSB (and as we will see GTSB) is in fact due to the shape and properties of the underlying Gallager regions. For this, only geometrical bodies with azimuthal symmetry along this radial vector are sought (as in the aforementioned bounds). In this fashion, the spherical symmetry of the Gaussian noise would lend itself to the simplicity of the calculation of the bound.

Definition: A *surface of revolution* is a surface generated by rotating a 2-dimensional curve about an axis [27].

Examples of surfaces of revolution include cone, cylinder, hyperboloid, paraboloid, and sphere. The important characteristic trait of all these-consequential in our work-is their azimuthal symmetry [20]. This translates to having sphere cross sections along the symmetry axis. This last property can be easily extended algebraically to higher dimensions.

In n-dimensional space, the following expression algebraically describes a body with azimuthal symmetry along the x_n axis in Cartesian coordinates:

$$x_1^2 + x_2^2 + \cdots + x_{n-1}^2 = r^2(x_n) \tag{5}$$

where $r(.)$ is an arbitrary function characterizing the cross sections [21].

The body defined in (5) will be referred to as a Hyper Surface of Revolution (HSR) whose azimuthal (symmetry/rotation) axis is x_n [28]. For a simple *right circular*[5] *n-cone*, with x_n as its axis and its apex at the origin, $r(x_n) = \alpha x_n$; where α is a constant scaling the solid angle of the n-cone[6] [17,22]. $r(x_n) = \alpha$ (α a constant) corresponds to a hyper-cylinder. A paraboloid is a surface of revolution of a parabola with the general $r(x_n) = \sqrt{|\alpha x_n|}$ (α a constant). Sphere accepts the formulation in (5) with

$$r(x_n) = \sqrt{\alpha^2 - x_n^2}, \qquad |x_n| \leq |\alpha|, \qquad \alpha \quad \text{a constant.}$$

[5]As opposed to other types of cones such as elliptic etc. For an elliptic cone, the cross sections are ellipses instead of spheres (circles). Right, as the apex is right above the center of the base.

[6]From this point on, by cone we mean a right circular cone.

Many of the other tight bounds developed for binary codes also use geometrical bodies fitting into the general framework of the (5) such as those in [15–17] which use conical or spherical regions.

V. Expansion of the Bound

From this point on, we consider the transmitted signal point, s_0, as the origin of the n-space. Separating the radial component of noise z_1 from the rest of the noise vector, one can expand the word error probability $P_w(E)$ as such:

$$P_w(E) = \int_{-\infty}^{+\infty} P(E|z_1) f_{z_1}(z_1) dz_1, \tag{6}$$

where $f_{z_1}(z_1)$ is the zero-mean Gaussian probability density function (pdf) with a variance of σ^2.

We choose the Gallager region $\Re$ to be an HSR with an azimuthal axis z_1 (see Fig. 3) and a general function $r(.)$ to be optimized shortly. Within this groundwork with a euclidean weight enumerating (ewe) function

$$ewe(w) = \sum_{j=1}^{n} A_j w^{\delta_j} \tag{7}$$

where A_j is the number of signal points at a Euclidean distance of $\delta_j = 2\sqrt{d_j}$ from s_0. Thus, we have:

$$P(E|z_1) \leq \min_{r(z_1)} \left\{ \sum_{k:\beta_k(z_1)<|r(z_1)|} A_k \cdot P(E_k|z_1, y \leq r^2(z_1)) + P(y > r^2(z_1)) \right\} \tag{8}$$

where $y = \sum_{i=2}^{n} z_i^2$ is a random variable with Chi-square distribution with $(n-1)$ degrees of freedom [29], i.e.,

$$f_y(y) = \frac{1}{2^{\frac{n-1}{2}} \Gamma(\frac{n-1}{2}) \sigma^{n-1}} \cdot e^{-\frac{y}{2\sigma^2}} y^{\frac{n-1}{2}-1} U(y) \tag{9}$$

where $\Gamma(.)$ and $U(.)$ are the complete gamma function and unit step function, respectively, E_k is the error event that the received vector r is closer to s_k (assuming $d_k = d(c_k, c_0)$), than the transmitted s_0, that is,

$$E_k = \{\|r - s_k\| \leq \|r - s_0\| | s_0\} \tag{10}$$

and $\beta_k(z_1)$, as seen in Fig. 3, is the projection of the perpendicular bisector hyper-plane between s_0 and s_k onto the $z_1 - z_2$ plane, that is, the straight line

$$\beta_k(z_1) = \frac{\sqrt{n} - z_1}{\sqrt{\frac{n}{d_k} - 1}}. \tag{11}$$

$\beta_k(z_1)$ is in fact the only entity in the development of the bound that solely applies to sphere constellations, hence, making the bound limited to the equi-energy signal sets.

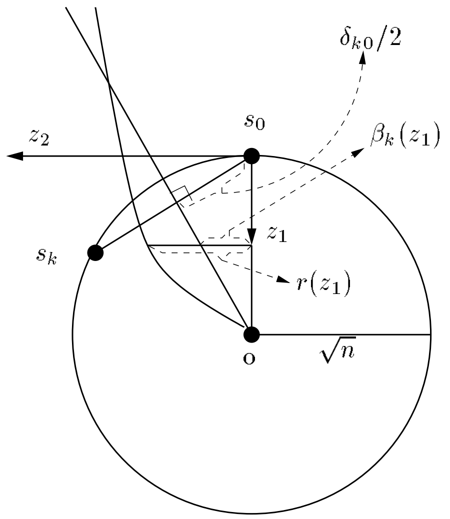

Fig. 3. Geometry of the general tangential sphere bound. The $z_1 - z_2$ plane is defined by the three points: origin (o), s_0, and s_k.

Now, by further separating the tangential component of noise z_2 ($z_2 \perp z_1$) from the complete noise vector we have,

$$P(E_k|z_1, y \leq r^2(z_1)) = P(\beta_k(z_1) < z_2 < |r(z_1)|, y_1 \leq r^2(z_1) - z_2^2) \tag{12}$$

where y_1 is a Chi-square distribution with $(n-2)$ degrees of freedom

$$f_{y_1}(y_1) = \frac{1}{2^{\frac{n-2}{2}}\Gamma(\frac{n-2}{2})\sigma^{n-2}} \cdot e^{-\frac{y_1}{2\sigma^2}} y_1^{\frac{n-2}{2}-1} U(y_1). \tag{13}$$

Therefore, the overall bound in (6) can be written as

$$\begin{aligned} P_w(E) \leq \min_{r(z_1)} \Bigg\{ \int_{-\infty}^{+\infty} \Big[\sum_{k:\beta_k(z_1)<|r(z_1)|} \Big(A_k \cdot \int_{\beta_k(z_1)}^{|r(z_1)|} f_{z_2}(z_2) \cdot \int_0^{r^2(z_1)-z_2^2} f_{y_1}(y_1) dy_1 \cdot dz_2 \Big) \\ + \int_{r^2(z_1)}^{+\infty} f_y(y) dy \Big] f_{z_1}(z_1) dz_1 \Bigg\} \end{aligned} \tag{14}$$

where z_2 as well as z_1 is a zero-mean Gaussian random variable with a variance σ^2.

Theorem 1: The bound in (14) is minimum for an $r(z_1)$ which is a linear function of $(z_1 - \sqrt{n})$.

Proof: Defining:

$$F[r(z_1)] = \int_{-\infty}^{+\infty} \Big[\sum_{k:\beta_k(z_1)<|r(z_1)|} \Big(A_k \cdot \int_{\beta_k(z_1)}^{|r(z_1)|} f_{z_2}(z_2) \cdot \int_{0}^{r^2(z_1)-z_2^2} f_{y_1}(y_1) dy_1 \cdot dz_2 \Big) + \int_{r^2(z_1)}^{+\infty} f_y(y) dy \Big] f_{z_1}(z_1) dz_1 \tag{15}$$

the functional in (15) will yield a stationary point if $\partial F[r(z_1) + \epsilon h(z_1)]/\partial\epsilon|_{\epsilon=0}$ is zero for all choices of $h(z_1)$ [30]. Using [31]

$$\frac{d}{dx} \int_{u(x)}^{v(x)} f(t,x) dt = f(v(x), x) \frac{dv}{dx} - f(u(x), x) \frac{du}{dx} + \int_{u(x)}^{v(x)} \frac{\partial}{\partial x} f(t,x) dt \tag{16}$$

and straightforward manipulations we end up with the equation:

$$\int_{-\infty}^{+\infty} \Big(\frac{1}{\sqrt{\pi}\Gamma(\frac{n-2}{2})} \cdot \sum_{k:\beta_k(z_1)<|r(z_1)|} \Big\{ A_k \cdot \int_{\beta_k(z_1)}^{|r(z_1)|} (r^2(z_1) - z_2^2)^{\frac{n}{2}-2} dz_2 \Big\} - \frac{1}{\Gamma(\frac{n-1}{2})} \cdot r^{n-3}(z_1) \Big) \cdot \Big[2r(z_1) \exp\left(-\frac{r^2(z_1)}{2\sigma^2} \right) h(z_1) f_{z_1}(z_1) \Big] dz_1 = 0. \tag{17}$$

By a change of variable in the integral over z_2, the above can be further simplified to:

$$\int_{-\infty}^{+\infty} \Big(\frac{1}{\sqrt{\pi}\Gamma(\frac{n-2}{2})} \cdot \sum_{k:\beta_k(z_1)<|r(z_1)|} \Big\{ A_k \cdot \int_{0}^{\cos^{-1}(\frac{\beta_k(z_1)}{|r(z_1)|})} \sin^{n-3}\theta d\theta \Big\} - \frac{1}{\Gamma\left(\frac{n-1}{2}\right)} \Big) \cdot \Big[2r^{n-2}(z_1) \exp\left(-\frac{r^2(z_1)}{2\sigma^2} \right) h(z_1) f_{z_1}(z_1) \Big] dz_1 = 0. \tag{18}$$

The equation in (18) will be satisfied for all $h(z_1)$ if the fraction $\frac{\beta_k(z_1)}{|r(z_1)|}$ is independent of z_1. This will require the function $r(z_1)$ to have the linear form $r(z_1) = r_0(z_1 - \sqrt{n})$ (where r_0 is a constant) and the optimization equation will reduce to:

$$\sum_{k:d_k \le \lfloor \frac{r_0^2 n}{1+r_0^2} \rfloor} A_k \cdot \int_0^{\theta_k} \sin^{n-3}\theta d\theta = \frac{\sqrt{\pi}\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \tag{19}$$

where

$$\theta_k = \cos^{-1}\left(\sqrt{\frac{d_k}{r_0^2(n-d_k)}}\right). \tag{20}$$

■

In other words, the optimum Gallager region is a cone whose apex is at the origin and its main axis is along the radial component of the noise. It should be noted that (19) corresponds to the result of Poltyrev [16]. Also, the summation upper limit in the optimization equation (19) is only valid for the upper nappe of the cone. For the lower nappe, $\beta_k(z_1)$ is negative and, therefore, the inequality $\beta_k(z_1) < 0 < |r(z_1)|$ is satisfied by all existing Hamming weights of the code from 1 to n. As the lower nappe probability has only marginal effect on the total error probability, the optimization in (19) will be sufficient for all values of z_1.

For linear binary block codes, this is a mathematical proof for what intuition would suggest. The Voronoi region of a transmitted codeword for a BPSK modulation binary code is the region surrounded by at most $(2^k - 1)$ hyper-planes all going through the origin at least $2\sqrt{d_{min}}$ apart from the communicated point of the constellation. This is a polyhedral cone with a single vertex at the origin of the n-space and unboundedly extending in one (radial) direction. This provides an intuitive explanation as to why the optimum Gallager region is a cone. This observation is not as straightforward as it may seem for non-binary codes. We emphasize that the application of the GTSB is not limited by any means to linear codes or binary alphabets or GU constellations. In fact, the only property of the scheme necessary for the bound is its being equi-energy (sphere code) which is required to keep the value of $\beta_k(z_1)$ valid as provided in (11). The bound in (14) applies to any sphere constellation with a given $ewe(w)$ function (which may depend on the center under consideration). This includes codes over non-binary alphabets, such as MPSK BCM schemes, as well as those which are not GU. In the latter, while the signal set is still equi-energy, one can use the proposed method to evaluate the error probability given a particular transmitted signal point; provided that the Euclidean spectrum centered at that point is available.

VI. Conclusions

Tightening of the caps on the ML decoding error probability of binary codes from the TB of Berlekamp to the TSB of Poltyrev has been an evolutionary process: an evolution of the Gallager region from a half-space in the TB to a sphere in the SB and finally to a cone in TSB, essentially, closing the gap between the Gallager and Voronoi regions.

The question still remained whether or not changing the boundaries of the Gallager region from a first-order function of the radial component of noise to any other function would be beneficiary to the tightness of the cap. This work extends the aforementioned boundary to any general one and proves that for a sphere code one cannot do better than a cone. This terminates the search for a better Gallager region and therefore a tighter bound within the format of the GTSB/TSB. The proposed bound and the method of proof for the optimality of the underlying Gallager region are also applicable to signal sets over non-binary alphabets and in general to any signal set as long as the signal points are all of equal energy. The proposed method can also be used to provide tight upper bounds in other applications such as mismatched decoding [32, 33] and nonuniform signaling [2] (a work under preparation by the authors).

Acknowledgments

The authors would like to thank K. Narimani, Prof. E.-H. Yang, and Prof. E. Vrscay for helpful discussions during the course of this work.

References

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

[2] H. Kuai, F. Alajaji, and G. Takahara, "Tight error bounds for nonuniform signaling over AWGN Channels," *IEEE Trans. Inform. Theory*, vol. IT-46, no. 7, pp. 2712-2718, Nov. 2000.

[3] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error correcting coding and decoding: Turbo codes," in *Proc. 1993 IEEE Int. Conf. on Communications,* Geneva, Switzerland, pp. 1064-1070, 1993.

[4] D. Divsalar, H. Jin, and R. J. McEliece, " Coding theorems for 'Turbo-like' codes," 1998 Allerton Conference, Sept. 23-25, 1998.

[5] R. G. Gallager, *Low Density parity Check Codes,* Cambridge, MA: MIT Press, 1963.

[6] D. J. C. MacKay and R. M. Neal, "Near Shannon limit performance of low density parity check codes," *Electron. Lett.,* vol. 32, pp. 1645-1646, Aug. 1996.

[7] R. G. Gallager, *Information Theory and Reliable Communication,* New York, NY: John Wiley & Sons, 1968.

[8] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups,* New York, NY: Springer-Verlag, 1988.

[9] 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.

[10] 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.

[11] 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.

[12] D. Slepian, "A class of binary signaling alphabets," *Bell Sys. Tech. Journ.,* vol. 35, pp. 203-234, Jan. 1956.

[13] D. Slepian, "Group codes for the Gaussian channel," *Bell Syst. Tech. J.,* vol. 47, no. 4, pp. 572-602, Apr. 1968.

[14] G. D. Forney, Jr., "Geometrically uniform codes," *IEEE Trans. Inform. Theory*, vol. IT-37, no. 5, pp. 1241-1260, Sept. 1991.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] E. R. Berlekamp, "The technology of error-correcting codes," *Proc. IEEE,* vol. 68, pp. 564-593, May 1980.

[20] D. Hilbert and S. Cohn-Vossen, *Geometry and the Imagination,* New York, NY: Chelsea, 1952.

[21] E. W. Weisstein, Eric Weisstein's World of Mathematics [Online]. Available: http://mathworld.wolfram.com.

[22] I. Sason and S. Shamai(Shitz), "Improved upper bounds on the decoding error probability of parallel and serial concatenated Turbo codes via their ensemble distance spectrum," *IEEE Trans. Inform. Theory*, vol. IT-46, no. 1, pp. 1-23, Jan. 2000.

[23] —, "Improved upper bounds on the ensemble performance of ML decoded low density parallel check codes," *IEEE Commun. Lett.,* vol. 4, pp. 89-91, Mar. 2000.

[24] —, "On improved bounds on the decoding error probability of block codes over interleaved fading channels, with applications to Turbo-like codes," *IEEE Trans. Inform. Theory*, vol. IT-47, no. 6, pp. 2275-2299, Nov. 2001.

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

[26] E. Agrell, "Voronoi regions for binary linear block codes," *IEEE Trans. Inform. Theory*, vol. IT-42, no. 1, pp. 310-316, Jan. 1996.

[27] D. Hilbert and S. Cohn-Vossen, *The Cylinder, the Cone, the Conic Sections, and Their Surfaces of Revolution*, New York, NY: Chelsea, 1999.

[28] P. M. Gruber and J. M. Wills, *Handbook of Convex Geometry,'* vol. A-B, Amsterdam, The Netherlands: North-Holland, 1993.

[29] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions,* vol. 1, New York, NY: John Wiley & Sons, 1994.

[30] I. M. Gelfand and S. V. Fomin, *Calculus of Variations,* Englewood Cliffs, NJ: Prentice-Hall, 1963.

[31] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, New York, NY: Academic Press, 1994.

[32] A. Ganti, A. Lapidoth, and I. E. Telatar, "Mismatched decoding revisited: general alphabets, channels with memory, and the wide-band limit," *IEEE Trans. Inform. Theory*, vol. IT-46, no. 7, pp. 2315-2328, Nov. 2000.

[33] N. Mehrav, G. Kaplan, A. Lapidoth, and S. Shamai, "On information rates for mismatched decoders," *IEEE Trans. Inform. Theory*, vol. IT-40, no. 6, pp. 1953-1967, Nov. 1994.