On the Error Exponents of Improved Tangential Sphere Bounds

Moshe Twitto, *Graduate Student Member, IEEE,* Igal Sason, *Member, IEEE*

Abstract

The performance of maximum-likelihood (ML) decoded binary linear block codes over the AWGN channel is addressed via the tangential sphere bound (TSB) and two of its recent improved versions. The paper is focused on the derivation of the error exponents of these bounds. Although it was exemplified that some recent improvements of the TSB tighten this bound for finite-length codes, it is demonstrated in this paper that their error exponents coincide. For an arbitrary ensemble of binary linear block codes, the common value of these error exponents is explicitly expressed in terms of the asymptotic growth rate of the average distance spectrum.

Index Terms

Block codes, bounds, linear codes, maximum-likelihood (ML) decoding.

I. Introduction

In recent years, much effort has been put into the derivation of tight performance bounds on the error probability of linear block codes under soft-decision maximum-likelihood (ML) decoding. During the last decade, this research work was stimulated by the introduction of various codes defined on graphs and iterative decoding algorithms, achieving reliable communication at rates close to capacity with feasible complexity. The remarkable performance of these codes at rates above the channel cut-off rate makes the union bound useless at a portion of the rate region where their performance is most appealing. Hence, tighter performance bounds are required to gain some insight on the performance of these efficient codes. Improved upper and lower bounds on the error probability of linear codes under ML decoding are addressed in [12] and references therein, and these bounds are applied to various codes and ensembles.

The tangential sphere bound (TSB) [9] forms one of the tightest upper bounds on the error probability for ML decoded linear block codes whose transmission takes place over the binary-input additive white Gaussian noise (BIAWGN) channel. The TSB was modified by Sason and Shamai [10] for the analysis of the bit error probability of linear block codes, and was slightly refined by Zangl and Herzog [19]. This bound only depends on the distance spectrum of the code, and hence, it can be applied to various codes and ensembles. The TSB falls within the class of upper bounds whose derivation relies on the basic inequality

$$\begin{aligned}\Pr(\text{word error} \mid \mathbf{c}_0) \leq & \Pr(\text{word error},\ \mathbf{y} \in \mathcal{R} \mid \mathbf{c}_0) \\ & + \Pr(\mathbf{y} \notin \mathcal{R} \mid \mathbf{c}_0),\end{aligned} \tag{1}$$

where $\mathbf{c}_0$ is the transmitted codeword, $\mathbf{y}$ denotes the received vector at the output of the channel, and $\mathcal{R}$ designates an arbitrary geometrical region which can be interpreted as a subset of the observation space. The idea is to use the union bound only for the joint event where the decoder fails to decode correctly and the received vector falls inside the region $\mathcal{R}$ (i.e., the union bound is used for upper bounding the first term on the right-hand side (RHS) of (1)). The TSB, for example, uses a circular hyper-cone as the region $\mathcal{R}$. Other upper bounds from this family are addressed in [12, Sections 3 and 4], [18] and references therein. In [15], Yousefi and Khandani prove that among all the volumes $\mathcal{R}$ which possess some symmetry properties, the circular hyper-cone yields the tightest bound. This finding demonstrates the optimality of the TSB among a family of bounds associated with geometrical regions

The paper was submitted in March 2006, and it was revised in November 2006. It was presented in part in the 2006 IEEE 24th Convention of Electrical and Electronics Engineers in Israel, November 15–17, Eilat, Israel.

The authors are with the Department of Electrical Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel (email: {tmoshe@tx, sason@ee}.technion.ac.il).

Communicated by R. J. McEliece, Associate Editor for Coding Theory.

which possess some symmetry properties, and which are obtained by applying the *union bound* to the first term on the RHS of (1). In [16], Yousefi and Khandani suggest to use Hunter's bound [8] (an upper bound which belongs to the family of second order Bonferroni-type inequalities [5]) instead of the union bound. This modification results in a tighter upper bound, referred to as the added hyper plane (AHP) bound. Yousefi and Mehrabian [17] also rely on Hunter's bound, but implement it in a quite different way in order to obtain an improved tangential sphere bound (ITSB) which solely depends on the distance spectrum of the code. The tightness of the ITSB and the AHP bound is exemplified in [16] and [17] for some binary linear block codes of short block lengths; these bounds slightly outperform the TSB at low SNR values.

An issue which is not addressed analytically in [16] and [17] is whether the new upper bounds (namely, the AHP bound and the ITSB) provide an improved lower bound on the error exponent as compared to the error exponent of the TSB. In this paper, we address this question, and prove that the error exponents of these improved bounds coincide with the error exponent of the TSB. We note however that the TSB fails to reproduce the random coding error exponent, especially for high-rate linear block codes [9].

This correspondence is organized as follows: The TSB ([9], [10]), the AHP bound [16] and the ITSB [17] are presented as a preliminary material in Section II (some boundary effects, which are not considered in [16] and [17], are discussed in detail in Section II). In Section III, we derive the error exponents of the ITSB and the AHP bound, and state our main result. We conclude our discussion in Section IV. Appendices provide supplementary details related to the proof of our main result.

II. Preliminaries

We introduce in this section some preliminary material which serves as a preparatory step towards the presentation of the material in the following section. We also present notation from [1] which is useful for our analysis. The reader is referred to [12] and [18] which introduce material covered in this section. However, in the following presentation, we consider boundary effects which were not taken into account in the original derivation of the TSB and its recent two improved versions in [7], [9], [16]–[18]. These boundary effects do not have any implication on the asymptotic exponential behavior of these bounds where we let the block length of the codes tend to infinity.

A. Assumption

Throughout this paper, we assume a binary-input additive white Gaussian noise (AWGN) channel with double-sided power spectral density of $\frac{N_0}{2}$. The modulation of the transmitted signals is antipodal, and the modulated signals are coherently detected and ML decoded (with soft decision).

B. Tangential Sphere Bound

The TSB forms an upper bound on the error probability of linear block codes under ML decoding where the transmission takes place over a binary-input AWGN channel (see [9]–[12]). Consider an (n, k) binary linear block code $\mathcal{C}$ of rate $R \triangleq \frac{k}{n}$ bits per channel use. Let us designate the codewords of $\mathcal{C}$ by $\{\mathbf{c}_i\}$ where $i = 0, 1, \ldots, 2^k - 1$. Assume a BPSK modulation, and let $\mathbf{s}_i = (s_{i,1}, \ldots, s_{i,n}) \in \{+\sqrt{E_\mathrm{s}}, -\sqrt{E_\mathrm{s}}\}^n$ designate the corresponding equi-energy modulated vectors where E_s denotes the energy per symbol. The transmitted signal vectors $\{\mathbf{s}_i\}$ are obtained from the codewords $\{\mathbf{c}_i\}$ by the mapping $s_{i,j} = (-1)^{c_{i,j}}\sqrt{E_\mathrm{s}}$ where $j = 1, 2, \ldots, n$, so the common value of their energy is equal to nE_s. Since the BIAWGN channel is memoryless and output-symmetric, and the code is linear, then the conditional error probability under ML decoding does not depend on the transmitted codeword. Hence, without any loss of generality, one can assume that the all-zero codeword $(\mathbf{c}_0)$ is transmitted; this corresponds to the signal vector $\mathbf{s}_0 = (+\sqrt{E_\mathrm{s}}, \ldots, +\sqrt{E_\mathrm{s}})$. Under this assumption, the received vector $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ is given by

$$y_j = \sqrt{E_\mathrm{s}} + z_j \ , \quad j = 1, 2, \ldots, n \tag{2}$$

where $\mathbf{z} = (z_1, z_2, \ldots, z_n)$ designates an n-dimensional Gaussian noise vector which corresponds to n orthogonal projections of the AWGN. Since $\mathbf{z}$ is a Gaussian vector and all its components are un-correlated, then the n components of $\mathbf{z}$ are statistically independent and identically distributed (i.i.d.); each component of $\mathbf{z}$ has a zero mean and variance $\sigma^2 = \frac{N_0}{2}$.

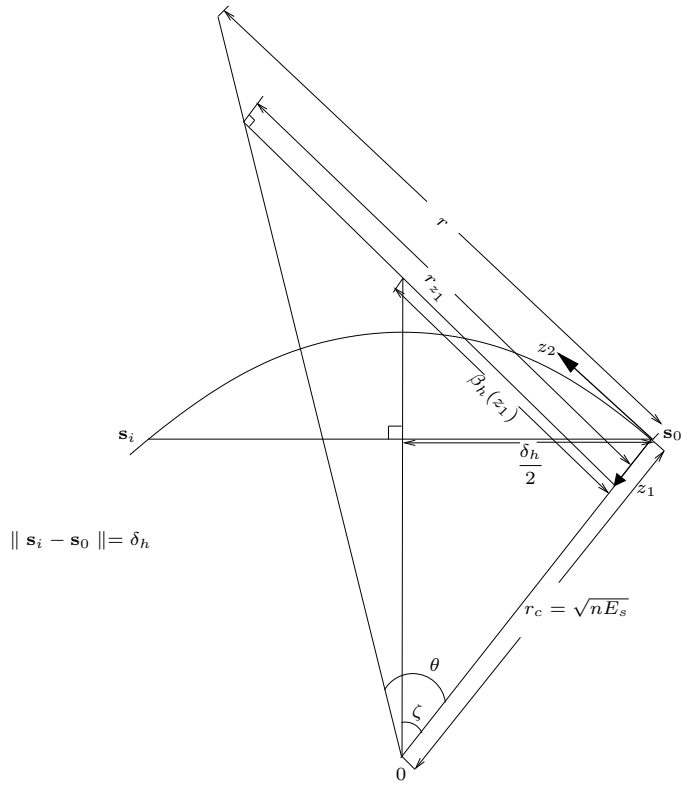

Fig. 1. The geometric interpretation of the TSB.

Let E be the event of deciding erroneously (under ML decoding) on a codeword other than the transmitted one. The TSB is based on the inequality

$$\Pr(E \mid \mathbf{c}_0) \leq \Pr(E,\ \mathbf{y} \in \mathcal{R} \mid \mathbf{c}_0) + \Pr(\mathbf{y} \notin \mathcal{R} \mid \mathbf{c}_0) \tag{3}$$

where $\mathcal{R}$ is an n-dimensional circular cone with a half angle θ whose vertex is located at the origin and its central line passes through the origin and the signal vector $\mathbf{s}_0$ (see Fig. 1). Let us designate this circular cone by $C_n(\theta)$.

The optimization of the upper bound on the decoding error probability (3) with $\mathcal{R} = C_n(\theta)$ is carried over r (where r and θ are related as shown in Fig. 1). Let z_1 be the radial component of the noise vector $\mathbf{z}$ (see Fig. 1), so the other $n-1$ components of $\mathbf{z}$ are orthogonal to the radial component z_1. From Fig. 1, we obtain that

$$r = \sqrt{nE_{\mathrm{s}}} \tan\theta, \quad r_{z_1} = \left(\sqrt{nE_{\mathrm{s}}} - z_1\right) \tan\theta$$

$$\beta_h(z_1) = \left(\sqrt{nE_{\mathrm{s}}} - z_1\right) \tan\zeta = \frac{\sqrt{nE_{\mathrm{s}}} - z_1}{\sqrt{nE_{\mathrm{s}} - \frac{\delta_h^2}{4}}} \frac{\delta_h}{2} \tag{4}$$

where $\delta_h = 2\sqrt{hE_{\mathrm{s}}}$ is the Euclidean distance between two BPSK modulated signal vectors, $\mathbf{s}_0$ and $\mathbf{s}_i$, which correspond to two codewords ($\mathbf{c}_0$ and $\mathbf{c}_i$, respectively) whose Hamming distance is h. The random variable

$$Y \triangleq \sum_{i=2}^{n} Z_i^2$$

is χ^2 distributed with $n-1$ degrees of freedom, so its *pdf* is given by

$$f_Y(y) = \frac{y^{\frac{n-3}{2}} e^{-\frac{y}{2\sigma^2}} U(y)}{2^{\frac{n-1}{2}} \sigma^{n-1} \Gamma\left(\frac{n-1}{2}\right)} \tag{5}$$

where the function U designates the unit step function, and the function Γ is the complete Gamma function

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt, \quad \mathrm{Re}(x) > 0. \tag{6}$$

Conditioned on the value of the radial component of the noise, z_1, let $E(z_1)$ designate the decoding error event. The conditional error probability satisfies the inequality

$$\begin{aligned}\Pr(E(z_1) \mid z_1) \leq & \Pr\left(E(z_1), \mathbf{y} \in C_n(\theta) \mid z_1\right) \\ & + \Pr\left(\mathbf{y} \notin C_n(\theta) \mid z_1\right). \end{aligned} \tag{7}$$

The conditional error event, $E(z_1)$, can be expressed as a union of pairwise error events, so

$$\begin{aligned}&\Pr(E(z_1), \mathbf{y} \in C_n(\theta) \mid z_1) \\ &= \Pr\left(\bigcup_{i=1}^{M-1} E_{0,i}(z_1), \mathbf{y} \in C_n(\theta) \mid z_1\right), \qquad M \triangleq 2^k\end{aligned} \tag{8}$$

where $E_{0,i}(z_1)$ designates the error event had the only codewords been $\mathbf{c}_0$ and $\mathbf{c}_i$, given the value z_1 of the radial component noise in Fig. 1; $M \triangleq 2^k$ denotes the number of codewords of the code $\mathcal{C}$. We note that for BPSK modulation, the Euclidean distance between the two signals $\mathbf{s}_i$ and $\mathbf{s}_0$ is directly linked to the Hamming weight of the codeword $\mathbf{c}_i$. Let the Hamming weight of $\mathbf{c}_i$ be h, then the Euclidean distance between $\mathbf{s}_0$ and $\mathbf{s}_i$ is equal to $\delta_h = 2\sqrt{hE_{\mathrm{s}}}$. Let $\{A_h\}$ be the distance spectrum of the linear code $\mathcal{C}$, and let $E_h(z_1)$ be the event of deciding under ML decoding in favor of other codeword $\mathbf{c}_i$ whose Hamming weight is h, given the value of z_1. By applying the union bound on the RHS of (8), we get

$$\begin{aligned}&\Pr(E(z_1), \mathbf{y} \in C_n(\theta) \mid z_1) \\ &\leq \sum_{h=1}^{n} A_h \Pr(E_h(z_1), \mathbf{y} \in C_n(\theta) \mid z_1).\end{aligned} \tag{9}$$

Combining (7) and (9) gives

$$\begin{aligned}\Pr\left(E(z_1) \mid z_1\right) \leq & \sum_h \Big\{ A_h \Pr\left(E_h(z_1), \mathbf{y} \in C_n(\theta) \mid z_1\right) \Big\} \\ & + \Pr\left(\mathbf{y} \notin C_n(\theta) \mid z_1\right).\end{aligned} \tag{10}$$

The second term on the RHS of (10) is evaluated from (5)

$$\begin{aligned}\Pr(\mathbf{y} \notin C_n(\theta) \mid z_1) &= \Pr\left(Y \geq r_{z_1}^2 \mid z_1\right) \\ &= \int_{r_{z_1}^2}^{\infty} f_Y(y) dy \\ &= \int_{r_{z_1}^2}^{\infty} \frac{y^{\frac{n-3}{2}} e^{-\frac{y}{2\sigma^2}}}{2^{\frac{n-1}{2}} \sigma^{n-1} \Gamma\left(\frac{n-1}{2}\right)} dy.\end{aligned} \tag{11}$$

This integral can be expressed in terms of the incomplete Gamma function

$$\gamma(a, x) \triangleq \frac{1}{\Gamma(a)} \int_0^x t^{a-1} e^{-t} dt, \quad a > 0,\ x \geq 0 \tag{12}$$

and it is given by

$$\Pr(\mathbf{y} \notin C_n(\theta) \mid z_1) = 1 - \gamma\left(\frac{n-1}{2}, \frac{r_{z_1}^2}{2\sigma^2}\right). \tag{13}$$

Let z_2 designate the tangential component of the noise vector $\mathbf{z}$, which is on the plane that contains the signals $\mathbf{s}_0$, $\mathbf{s}_i$ and the origin, and orthogonal to z_1 (see Fig. 1). Referring to the first term on the RHS of (10), it follows from the geometry in Fig. 1 that if $z_1 \leq \sqrt{nE_{\mathrm{s}}}$ then

$$\begin{aligned}&\Pr(E_h(z_1), \mathbf{y} \in C_n(\theta) \mid z_1) \\ &= \Pr(E_h(z_1),\ Y \leq r_{z_1}^2 \mid z_1) \\ &= \Pr\left(\beta_h(z_1) \leq z_2 \leq r_{z_1},\ Y \leq r_{z_1}^2 \mid z_1\right).\end{aligned} \tag{14}$$

Let $V \triangleq \sum_{i=3}^{n} Z_i^2$, then $V = Y - Z_2^2$. If $z_1 \leq \sqrt{nE_{\rm s}}$, then we obtain the equality

$$\begin{aligned} &\Pr\left(E_h(z_1), \mathbf{y} \in C_n(\theta) \mid z_1\right) \\ &= \Pr\left(\beta_h(z_1) \leq Z_2 \leq r_{z_1}, V \leq r_{z_1}^2 - Z_2^2 \mid z_1\right). \end{aligned} \tag{15}$$

The random variable V is χ^2 distributed with $n-2$ degrees of freedom, so its *pdf* is

$$f_V(v) = \frac{v^{\frac{n-4}{2}} e^{-\frac{v}{2\sigma^2}} U(v)}{2^{\frac{n-2}{2}} \sigma^{n-2} \Gamma\left(\frac{n-2}{2}\right)} \tag{16}$$

and since the random variables V and Z_2 are statistically independent, then if $z_1 \leq \sqrt{nE_{\rm s}}$

$$\begin{aligned} &\Pr\left(E_h(z_1), \mathbf{y} \in C_n(\theta) \mid z_1\right) \\ &= \int_{\beta_h(z_1)}^{r_{z_1}} \frac{e^{-\frac{z_2^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} \int_0^{r_{z_1}^2 - z_2^2} f_V(v)\, dv\, dz_2. \end{aligned} \tag{17}$$

In order to obtain an upper bound on the decoding error probability, $\Pr(E)$, one should apply the statistical expectation operator on the RHS of (10) w.r.t. the radial noise component z_1. Referring to the upper half azimuthal cone depicted in Fig. 1, which corresponds to the case where the radial noise component satisfies the condition $z_1 \leq \sqrt{nE_{\rm s}}$, it follows from (4) that the inequality $\beta_h(z_1) < r_{z_1}$ holds for those values of h for which $\frac{\delta_h}{2} < \alpha_h$ where

$$\alpha_h \triangleq r\sqrt{1 - \frac{\delta_h^2}{4nE_{\rm s}}}. \tag{18}$$

From Fig. 1, if $z_1 > \sqrt{nE_{\rm s}}$, then the range of integration on the RHS of (17) for the noise component z_2 changes; as it corresponds to the joint event $(E_h(z_1), \mathbf{y} \in C_n(\theta))$, then the relevant interval of integration w.r.t. z_2 is $\beta_h(z_1) \leq z_2 \leq -r_{z_1}$. This inequality is meaningful for all values of h (since for $z_1 > \sqrt{nE_{\rm s}}$, we get from (4) that $r_{z_1} < 0$ and $\beta_h(z_1) < 0$, so the inequality $\beta_h(z_1) \leq -r_{z_1}$ holds in this case for all values of h). Since $Z_1 \sim N(0, \sigma^2)$ where $\sigma^2 = \frac{N_0}{2}$, then the probability that the Gaussian random variable Z_1 exceeds $\sqrt{nE_{\rm s}}$ is equal to

$$Q\left(\frac{\sqrt{nE_{\rm s}}}{\sigma}\right) = Q\left(\sqrt{\frac{2nRE_{\rm b}}{N_0}}\right)$$

where $E_{\rm b}$ is the energy per information bit, and $E_{\rm s} = RE_{\rm b}$.

This results in the following upper bound on the decoding error probability under ML decoding

$$\begin{aligned} \Pr(E) \leq &\int_{-\infty}^{+\sqrt{nE_{\rm s}}} \frac{e^{-\frac{z_1^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} \\ &\cdot \Bigg\{ \sum_{h:\frac{\delta_h}{2} < \alpha_h} \Bigg\{ A_h \int_{\beta_h(z_1)}^{r_{z_1}} \frac{e^{-\frac{z_2^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} \int_0^{r_{z_1}^2 - z_2^2} f_V(v)\, dv\, dz_2 \Bigg\} \\ &+ 1 - \gamma\left(\frac{n-1}{2}, \frac{r_{z_1}^2}{2\sigma^2}\right) \Bigg\} dz_1 + Q\left(\sqrt{\frac{2nRE_{\rm b}}{N_0}}\right). \end{aligned} \tag{19}$$

The upper bound on the error probability (19) is valid for all positive values of r. Hence, in order to achieve the tightest upper bound of the form (19), one sets to zero the partial derivative of the integrand on the RHS of (19) w.r.t. r_{z_1}; note that it follows from (4) that $r_{z_1} = \left(1 - \frac{z_1}{\sqrt{nE_s}}\right) r$, so for any value of $z_1 \leq \sqrt{nE_s}$, setting to zero the partial derivative of the integrand on the RHS of (19) w.r.t. r_{z_1} is equivalent to setting to zero its partial derivative w.r.t. r. Straightforward algebra gives the following optimization equation for the value of r [9]:

$$\begin{cases} \displaystyle\sum_{h:\frac{\delta_h}{2} < \alpha_h} A_h \int_0^{\theta_h} \sin^{n-3}\phi\, d\phi = \frac{\sqrt{\pi}\,\Gamma(\frac{n-2}{2})}{\Gamma(\frac{n-1}{2})} \\ \theta_h = \cos^{-1}\left(\frac{\delta_h}{2\alpha_h}\right) \end{cases} \tag{20}$$

where α_h is given in (18). A proof for the existence and uniqueness of a solution r to the optimization equation (20) is provided in [11, Appendix B], together with an efficient algorithm to solve this equation numerically. In order to derive an upper bound on the bit error probability, let $A_{w,h}$ designate the coefficient of the input-output weight enumeration function (IOWEF) which corresponds to the number of codewords that are encoded by information bits whose number of ones is equal to w (where $0 \leq w \leq nR$) and whose Hamming weight (after encoding) is equal to h, and let

$$A'_h \triangleq \sum_{w=1}^{nR} \left(\frac{w}{nR}\right) A_{w,h}, \quad h = 0, \ldots, n. \tag{21}$$

In [11, Appendix C], Sason and Shamai derive an upper bound on the bit error probability; the upper bound on the bit error probability is identical to the TSB on the block error probability, except of the replacement of the distance spectrum $\{A_h\}$ in (19) and (20) with the sequence $\{A'_h\}$, as in (21) (where $A'_h \leq A_h$ holds for all Hamming weights h, so the upper bound on the bit error probability is not larger than the upper bound on the block error probability, as expected).

C. Improved Tangential Sphere Bound

In [17], Yousefi and Mehrabian derive a new upper bound on the block error probability of binary linear block codes whose transmission takes place over a binary-input AWGN channel, and which are coherently detected and ML decoded. This upper bound, referred to as the improved tangential sphere bound (ITSB), originates from (3) where the corresponding region $\mathcal{R}$ on the RHS of (3) is the same as for the TSB (i.e., it is an n-dimensional circular cone whose main axis passes through the origin and the transmitted signal vector). However, unlike the TSB which relies on the union bound for bounding the first term on the RHS of (3), the derivation of the ITSB relies on Hunter's bound which is a Bonferroni-type inequality of the second order (see [5] and [8]); as compared to the union bound, the latter inequality is used in order to get a tighter upper bound on the probability of the joint event where there is a decoding error and the received vector falls within the n-dimensional circular cone $C_n(\theta)$ (see Fig. 1).

The basic idea in [17] relies on Hunter's bound which states that if $\{E_i\}_{i=1}^M$ designates a set of M events, and E_i^{c} designates the complementary event of E_i, then

$$\begin{aligned} \Pr\left(\bigcup_{i=1}^{M} E_i\right) &= \Pr(E_1) + \Pr(E_2 \cap E_1^c) \\ &\quad + \ldots + \Pr(E_M \cap E_{M-1}^{\mathrm{c}} \ldots \cap E_1^{\mathrm{c}}) \\ &\leq \Pr(E_1) + \sum_{i=2}^{M} \Pr(E_i \cap E_{\hat{i}}^{\mathrm{c}}). \end{aligned} \tag{22}$$

where the indices $\hat{i} \in \{1, 2, \ldots i-1\}$ are chosen arbitrarily for $i \in \{2, \ldots, M\}$. Clearly, the upper bound (22) is tighter than the union bound. The LHS of (22) is invariant to the ordering of the events (since it only depends on the union of these events) while the RHS of (22) depends on this ordering. Hence, the tightest bound of the form (22) is obtained by choosing the optimal indices ordering $i \in \{1, 2, \ldots, M\}$ and $\hat{i} \in \{1, 2, \ldots, i-1\}$. Let us designate by $\Pi(1, 2, \ldots, M) = \{\pi_1, \pi_2, \ldots, \pi_M\}$ an arbitrary permutation among the $M!$ possible permutations of the set $\{1, 2, \ldots, M\}$ (i.e., a permutation of the indices of the events E_1 to E_M), and let $\Lambda = (\lambda_2, \lambda_3, \ldots \lambda_M)$ designate an arbitrary sequence of integers where $\lambda_i \in \{\pi_1, \pi_2, \ldots \pi_{i-1}\}$. Then, the tightest form of of the bound in (22) is given by

$$\Pr\left(\bigcup_{i=1}^{M} E_i\right) \leq \min_{\Pi, \Lambda} \left\{ \Pr(E_{\pi_1}) + \sum_{i=2}^{M} \Pr(E_{\pi_i} \cap E_{\lambda_i}^c) \right\}. \tag{23}$$

Similar to the TSB, the derivation of the ITSB originates from the upper bound (7) on the conditional decoding error probability, given the radial component (z_1) of the noise vector (see Fig. 1). In [17], it is proposed to apply the upper bound (23) on the RHS of (8) which for an arbitrary permutation $\{\pi_1, \pi_2, \ldots, \pi_M\}$ and a corresponding

sequence of integers $(\lambda_2, \lambda_3, \ldots \lambda_{M-1})$ as above, gives

$$\begin{aligned}
&\Pr\left(\bigcup_{i=1}^{M-1} E_{0,i}(z_1), \mathbf{y} \in C_n(\theta) \mid z_1\right) \\
&\leq \min_{\Pi,\Lambda}\Bigg\{\Pr(E_{0,\pi_1}(z_1), \mathbf{y} \in C_n(\theta) \mid z_1) \\
&\qquad + \sum_{i=2}^{M-1} \Pr(E_{0,\pi_i}(z_1), E^{\mathrm{c}}_{0,\lambda_i}(z_1), \mathbf{y} \in C_n(\theta) \mid z_1)\Bigg\} \qquad (24)
\end{aligned}$$

where $E_{0,j}(z_1)$ designates the pairwise error event for which the decoder decides on codeword $\mathbf{c}_j$ rather than the transmitted codeword $\mathbf{c}_0$, given the radial component (Z_1) of the noise is equal to z_1. As indicated in [16] and [17], the optimization problem of (24) is prohibitively complex. In order to simplify it, Yousefi and Mehrabian suggest to choose $\pi_1 = \lambda_i = i_{\min}$ for all $i = 2, \ldots, M-1$, where $i_{\min}$ designates the index of a codeword which is closest (in terms of Euclidian distance) to the transmitted signal vector $\mathbf{s}_0$. Since the code is linear and the channel is memoryless and symmetric, one can assume without any loss of generality that the all-zero codeword is transmitted. Moreover, since we deal with antipodal modulation, then $w_{\mathrm{H}}(\mathbf{c}_{i_{\min}}) = d_{\min}$ where $d_{\min}$ is the minimum distance of the code. Hence, by this particular choice of π_1 and Λ (which in general loosen the bound in (24)), the ordering of the indices $\{\pi_2, \ldots, \pi_{M-1}\}$ is irrelevant, and one can omit the optimization over Π and Λ. The above simplification results in the following inequality:

$$\begin{aligned}
\Pr(E(z_1)|z_1) \leq &\Pr\left(E_{0,i_{\min}}(z_1), \mathbf{y} \in C_n(\theta) \mid z_1\right) \\
&+ \sum_{i=2}^{M-1} \Pr(E_{0,i}(z_1), E^{\mathrm{c}}_{0,i_{\min}}(z_1), \mathbf{y} \in C_n(\theta) \mid z_1) \\
&+ \Pr\left(\mathbf{y} \notin C_n(\theta) \mid z_1\right). \qquad (25)
\end{aligned}$$

Based on Fig. 1, the first term on the RHS of (25) satisfy the equality

$$\begin{aligned}
&\Pr\left(E_{0,i_{\min}}(z_1),\ \mathbf{y} \in C_n(\theta) \mid z_1\right) \\
&= \Pr(\beta_{\min}(z_1) \leq Z_2 \leq r_{z_1},\ V < r^2_{z_1} - Z_2^2) \qquad (26)
\end{aligned}$$

where it follows from (4)

$$\beta_{\min}(z_1) \triangleq \beta_{d_{\min}}(z_1) = \left(\sqrt{nE_{\mathrm{s}}} - z_1\right)\sqrt{\frac{d_{\min}}{n - d_{\min}}}. \qquad (27)$$

Recall that Z_2 is the tangential component of the noise vector $\mathbf{z}$ which is on the plane that contains the signals $\mathbf{s}_0$, $\mathbf{s}_{i_{\min}}$ and the origin (see Fig. 1), and the other parameters are introduced in (4). The third term on the RHS of (25) which corresponds to the probability that the received vector $\mathbf{y}$ falls outside the circular cone $C_n(\theta)$ is given in (13).

In order to express the probabilities of the form $\Pr(E_{0,i}(z_1), E^c_{0,i_{\min}}(z_1), \mathbf{y} \in C_n(\theta) \mid z_1)$ which are encountered on the RHS of (25), we rely on the geometry shown in the upper plot of Fig. 2. The BPSK modulated signals $\mathbf{s}_0$, $\mathbf{s}_i$ and $\mathbf{s}_j$, where $j = i_{\min}$, all lie on the surface of a hyper-sphere centered at the origin and with radius $\sqrt{nE_{\mathrm{s}}}$. The planes P_1 and P_2 are constructed by the triplets of points $(\mathbf{o}, \mathbf{s}_0, \mathbf{s}_i)$ and $(\mathbf{o}, \mathbf{s}_0, \mathbf{s}_j)$, respectively. In the derivation of the ITSB, Yousefi and Mehrabian choose $\mathbf{s}_j$ to correspond to codeword $\mathbf{c}_j$ with Hamming weight $d_{\min}$.

Let Z_3' be a noise component which is orthogonal to Z_1 and which lies on the plane P_2 (see the upper plot in Fig. 2). Based on the geometry shown in Fig. 2, if $Z_1 = z_1 \leq \sqrt{nE_s}$, then the joint event $(E_{0,i}(z_1), E^{\mathrm{c}}_{0,j}(z_1), \mathbf{y} \in C_n(\theta))$ is equivalent to the case where the projections of the received vector $\mathbf{y}$ on the planes P_1 and P_2 fall, respectively, inside the left and right dashed areas of this figure. One therefore obtains the following equality if $z_1 \leq \sqrt{nE_{\mathrm{s}}}$:

$$\begin{aligned}
&\Pr(E_{0,i}(z_1),\ E^c_{0,i_{\min}}(z_1),\ \mathbf{y} \in C_n(\theta) \mid z_1) \\
&= \Pr\Big(\beta_i(z_1) \leq Z_2 \leq r_{z_1}, \\
&\qquad - r_{z_1} \leq Z_3' \leq \beta_{\min}(z_1),\ Y < r^2_{z_1} \mid z_1\Big). \qquad (28)
\end{aligned}$$

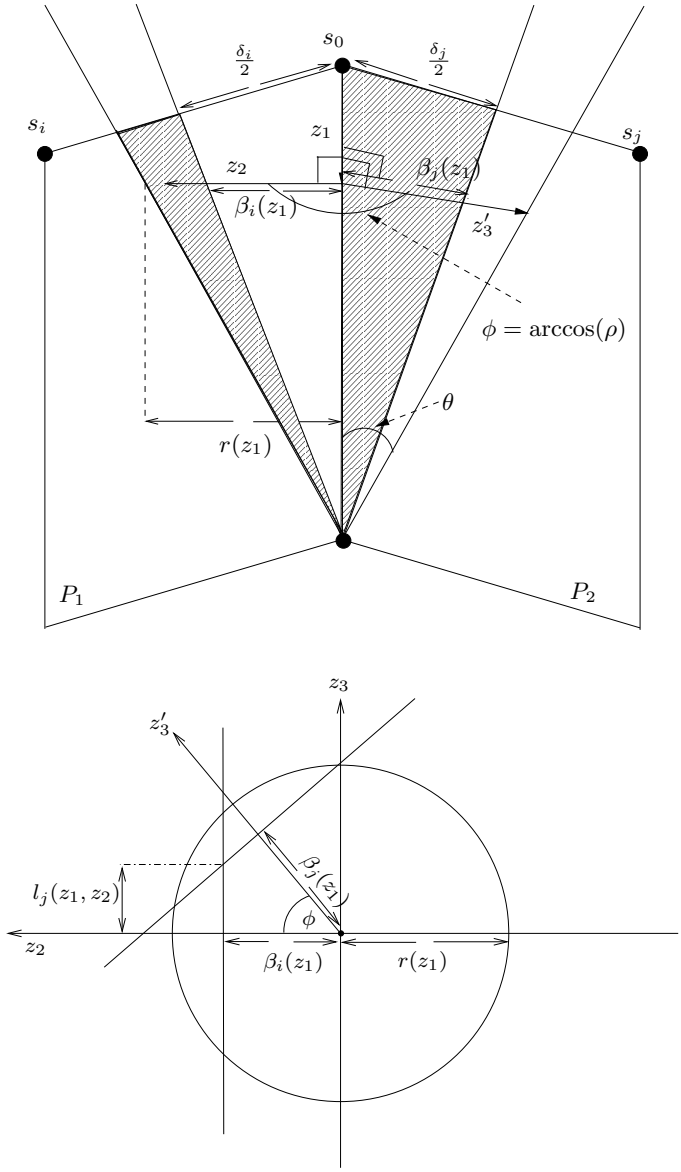

Fig. 2. Upper plot: $\mathbf{s}_0$ is the transmitted vector, z_1 is the radial noise component, z_2 and z_3' are two (not necessarily orthogonal) noise components, which are perpendicular to z_1, and lie on planes P_1 and P_2, respectively. The doted and dashed areas are the regions where the events $E_{0,i}$ and $E_{0,j}^c$ occur, respectively. Lower plot: A cross-section of the geometry in (a).

Furthermore, from the geometry shown in the lower plot of Fig. 2, it follows that

$$Z_3' = Z_3 \sin\phi + Z_2 \cos\phi \tag{29}$$

where Z_3 is a noise component which is orthogonal to the two noise components Z_1 and Z_2, and lies on the n-dimensional plane which is formed by the noise components Z_2 and Z_3' (see the lower plot of Fig. 2). Given that $Z_1 = z_1$, plugging (29) into the condition $-r_{z_1} \leq Z_3' \leq \beta_{\min}(z_1)$ in (28) yields that

$$-r_{z_1} \leq Z_3 \leq \min\Big\{ l(z_1, Z_2), r_{z_1} \Big\}$$

where

$$l(z_1, z_2) = \frac{\beta_{\min}(z_1) - \rho z_2}{\sqrt{1-\rho^2}} \tag{30}$$

and $\rho = \cos\phi$ is the correlation coefficient between the noise components whose sizes are z_2 and z_3' (see Fig. 2 where ϕ designates the angle between the planes P_1 and P_2).

Let $W=\sum_{i=4}^{n} Z_i^2$, then if $z_1 \le \sqrt{nE_{\rm s}}$, the following equality holds:

$$\begin{aligned}&\Pr(E_{0,i}(z_1),\ E^c_{0,i_{\min}}(z_1),\ \mathbf{y}\in C_n(\theta)\mid z_1)\\ &=\Pr\Bigg(\beta_i(z_1)\le Z_2\le r_{z_1},\ -r_{z_1}\le Z_3\le \min\Big\{l(z_1,Z_2), r_{z_1}\Big\},\\ &\qquad W<r_{z_1}^2-Z_2^2-Z_3^2\mid z_1\Bigg).\end{aligned} \tag{31}$$

The random variable W is Chi-squared distributed with $n-3$ degrees of freedom, so its *pdf* is given by

$$f_W(w)=\frac{w^{\frac{n-5}{2}}e^{-\frac{w}{2\sigma^2}}U(w)}{2^{\frac{n-3}{2}}\sigma^{n-3}\Gamma\left(\frac{n-3}{2}\right)}. \tag{32}$$

Since the probability

$$\Pr(E_{0,i}(z_1),\ E^c_{0,i_{\min}}(z_1),\ \mathbf{y}\in C_n(\theta)\mid z_1)$$

depends on the correlation coefficient between the noise components z_2 and z_3' (see Fig. 2), then the distance spectrum of the code does not provide sufficient information for the calculation of the bound. To circumvent this problem and obtain an upper bound which solely depends on the distance spectrum of the code, it is suggested in [17] to loosen the bound as follows. It is shown in [16, Appendix B] that while referring to two codewords of Hamming weights d_i and d_j (other than the transmitted codeword), the corresponding correlation coefficient as defined above (see Fig. 2) satisfies

$$\begin{aligned}&-\min\left\{\sqrt{\frac{d_id_j}{(n-d_i)(n-d_j)}},\sqrt{\frac{(n-d_i)(n-d_j)}{d_id_j}}\right\}\\ &\le\rho\le\frac{\min(d_i,d_j)[n-\max(d_i,d_j)]}{\sqrt{d_id_j(n-d_i)(n-d_j)}}.\end{aligned} \tag{33}$$

Moreover, the RHS of (31) is shown to be a monotonic decreasing function of ρ (see [17, Appendix 1] and a further discussion in Appendix III here). Hence, one can omit the dependency in the geometry of the code (and loosen the upper bound) by replacing the correlation coefficients in (31) with their lower bounds which solely depend on the weights of the two codewords. In the derivation of the ITSB, we consider the above correlation coefficients in Fig. 2 while referring to two codewords of Hamming weights $d_i=h$ $(h\le n)$ and $d_j=d_{\min}$ (other than the transmitted codeword). Let

$$\begin{aligned}\rho_h\ &\triangleq -\min\left\{\sqrt{\tfrac{hd_{\min}}{(n-h)(n-d_{\min})}},\sqrt{\tfrac{(n-h)(n-d_{\min})}{hd_{\min}}}\right\}\\ &=\begin{cases}-\sqrt{\frac{hd_{\min}}{(n-h)(n-d_{\min})}} & \text{if } d_{\min}+h\le n\\ -\sqrt{\frac{(n-h)(n-d_{\min})}{hd_{\min}}} & \text{if } d_{\min}+h>n\end{cases}.\end{aligned} \tag{34}$$

From (25) and (26) and by averaging w.r.t. Z_1, one gets the following upper bound on the decoding error probability:

$$\begin{aligned}&\Pr(E)\\ &\le\Pr\left(Z_1\le\sqrt{nE_{\rm s}},\ \beta_{\min}(Z_1)\le Z_2\le r_{Z_1}, V\le r_{Z_1}^2-Z_2^2\right)\\ &\quad+\sum_{h=d_{\min}}^{n}\Bigg\{A_h\Pr\Big(Z_1\le\sqrt{nE_{\rm s}},\ \beta_h(Z_1)\le Z_2\le r_{Z_1},\\ &\qquad\qquad -r_{Z_1}\le Z_3\le\min\Big\{l_h(Z_1,Z_2),\ r_{Z_1}\Big\},\end{aligned}$$

$$W \leq r_{Z_1}^2 - Z_2^2 - Z_3^2\Big)\Big\}$$
$$+\Pr\left(Z_1 \leq \sqrt{nE_{\rm s}},\ Y \geq r_{Z_1}^2\right) + \Pr(Z_1 > \sqrt{nE_{\rm s}}) \tag{35}$$

where the parameter $l_h(z_1, z_2)$ is simply $l(z_1, z_2)$ in (30) with the correlation ρ replaced by ρ_h in (34), i.e.,

$$l_h(z_1, z_2) \triangleq \frac{\beta_{\min}(z_1) - \rho_h z_2}{\sqrt{1-\rho_h^2}}. \tag{36}$$

Using the probability density functions of the random variables in the RHS of (35), and since the random variables Z_1, Z_2, Z_3 and W are statistically independent, the final form of the ITSB is given by

$$\begin{aligned}\Pr(E) \leq \int_{-\infty}^{\sqrt{nE_{\rm s}}} \Bigg[&\int_{\beta_{\min}}^{r_{z_1}} f_{Z_2}(z_2) \int_0^{r_{z_1}^2 - z_2^2} f_V(v) dv \cdot dz_2 \\ &+ \sum_{h:\, \beta_h(z_1) < r_{z_1}} \Bigg(A_h \int_{\beta_h(z_1)}^{r_{z_1}} \int_{-r_{z_1}}^{\min\{l_h(z_1,z_2), r_{z_1}\}} f_{Z_2,Z_3}(z_2, z_3) \\ &\qquad \int_0^{r_{z_1}^2 - z_2^2 - z_3^2} f_W(w)\, dw\, dz_2\, dz_3 \Bigg) \\ &+ 1 - \gamma\left(\frac{n-1}{2}, \frac{r_{z_1}^2}{2\sigma^2}\right)\Bigg] f_{Z_1}(z_1) dz_1 + Q\left(\sqrt{\frac{2nRE_{\rm b}}{N_0}}\right).\end{aligned} \tag{37}$$

Note that $V \triangleq \sum_{i=3}^{n} z_i^2$ and $W \triangleq \sum_{i=4}^{n} z_i^2$ are Chi-squared distributed with $(n-2)$ and $(n-3)$ degrees of freedom, respectively (see (16) and (32) for their probability density functions f_V and f_W, respectively).

D. Added-Hyper-Plane (AHP) Bound

In [16], Yousefi and Khandani introduce a new upper bound on the ML decoding block error probability, referred to as the added hyper plane (AHP) bound. Similar to the derivation of the ITSB, the AHP bound is based on Hunter's bound (22) which gives the inequality (24). The complicated optimization problem in (24), however, is treated differently.

Let us denote by $\mathcal{I}_w$ the set of the indices of the codewords of $\mathcal{C}$ with Hamming weight w. For $i \in \{1, 2, \ldots, M\} \backslash \mathcal{I}_w$, let $\{j_i\}$ be a sequence of integers chosen from the set $\mathcal{I}_w$. Then, based on (8) and the concept of Hunter's bound in (22), the following upper bound holds

$$\begin{aligned}&\Pr\left(E(z_1),\ \mathbf{y} \in C_n(\theta) \mid z_1\right) \\ &\leq \min_{w, \mathcal{J}_w} \Bigg\{ \Pr\Bigg(\bigcup_{j \in \mathcal{I}_w} \left\{E_{0,j}(z_1)\right\},\ \mathbf{y} \in C_n(\theta) \mid z_1 \Bigg) \\ &\qquad + \sum_{i \in \{1,\ldots,M-1\} \backslash \mathcal{I}_w} \Pr\left(E_{0,i}(z_1), E_{0,j_i}^c(z_1),\ \mathbf{y} \in C_n(\theta) \mid z_1\right) \Bigg\}.\end{aligned} \tag{38}$$

From (28) and the lower plot in Fig. 2, it is clear that the probabilities which appear in the second term of the RHS of (38) depend on the corresponding correlation coefficients between the noise components z_2 and z_3'. Hence, in order to compute the upper bound (38), one has to know the geometrical characterization of the Voronoi regions of the codewords. To obtain an upper bound which only requires the knowledge of the distance spectrum of the code, Yousefi and Khandani [16] suggest to extend the codebook by adding all the remaining $\binom{n}{w} - A_w$ n-tuples with Hamming weight w (i.e., to generate an extended block code which contains in addition all the binary vectors of length n and Hamming weight w). Let us designate the new code by $\mathcal{C}_w$ and denote its codewords by $\mathbf{c}_i^w$ where $i \in \left\{0, 1, \ldots, M + \binom{n}{w} - A_w - 1\right\}$. The new codebook is not necessarily linear, and all possible correlation coefficients $\rho = \cos\phi$ (see Fig. 2) are available while referring to two codewords other than the transmitted codeword where one of these codewords is of Hamming weight i, where $i \in \{d_{\min}, \ldots n\}$, and the other is of

Hamming weight w. Thus, for each layer of the codebook, one can choose the *largest available correlation* ρ with respect to any possible n-tuple binary vector of Hamming weight w; this is due to the fact that the RHS of (38) is a monotonically decreasing function of ρ, as shown in [17], so by choosing the largest available correlation in Fig. 2 where one of the signal points can be any binary vector of Hamming weight w tightens the upper bound on the error probability. Now, one may find the optimum layer at which the codebook extension is done, i.e., find the optimum $w \in \{1, 2, \dots n\}$ which yields the tightest upper bound within this form. We note that the resulting upper bound is not proved to be uniformly tighter than the TSB, due to the extension of the code. The maximal correlation coefficient ρ between the noise components whose sizes are z_2 and z_3' (see Fig. 2) while referring to two codewords of Hamming weights d_i and d_j is introduced on the RHS of (33) (see [16]). Let us designate the maximal possible correlation coefficient in Fig. 2 which refer to two n-tuples with Hamming weights w and h by $\rho_{w,h}$, i.e., based on (33) we define

$$\rho_{w,h} \triangleq \frac{\min(h,w)[n-\max(h,w)]}{\sqrt{hw(n-h)(n-w)}} \ , \quad w \neq h. \tag{39}$$

By using the same bounding technique as of the ITSB, and replacing the correlation coefficients with their respective upper bounds, $\rho_{w,h}$, (38) gets the form

$$\begin{aligned}
&\Pr\left(E(z_1),\ \mathbf{y} \in C_n(\theta) \mid z_1\right) \\
&\leq \min_{w} \Bigg\{ \Pr\left(\bigcup_{j:w_{\mathrm{H}}(\mathbf{c}_j^w)=w} E_{0,j}(z_1),\ \mathbf{y} \in C_n(\theta) \mid z_1 \right) \\
&\qquad + \sum_{h \neq w} \Bigg\{ A_h \ \Pr\big(Y \leq r_{z_1}^2,\ Z_2 \geq \beta_h(z_1), \\
&\qquad\qquad Z_3 \leq l_{w,h}(z_1, Z_2)\big) \Bigg\} \Bigg\}
\end{aligned} \tag{40}$$

where based on the lower plot of Fig. 2 and the explanation above with the correlation ρ replaced by $\rho_{w,h}$, we get

$$l_{w,h}(z_1, z_2) = \frac{\beta_w(z_1) - \rho_{w,h} z_2}{\sqrt{1-\rho_{w,h}^2}} \tag{41}$$

and $\rho_{w,h}$ is introduced in (39). Applying Hunter's bound to the first term on the RHS of (40) gives

$$\begin{aligned}
&\Pr\left(\bigcup_{j:w_{\mathrm{H}}(\mathbf{c}_j^w)=w} E_{0,j}(z_1),\ \mathbf{y} \in C_n(\theta) \mid z_1 \right) \\
&\leq \Pr(E_{0,l_0}(z_1),\ \mathbf{y} \in C_n(\theta) \mid z_1) \\
&\quad + \sum_{i=1}^{\binom{n}{w}-1} \Pr(E_{0,l_i}(z_1),\ E_{0,\hat{l}_i}^c(z_1),\ \mathbf{y} \in C_n(\theta) \mid z_1)
\end{aligned} \tag{42}$$

where $\{l_i\}$ for $i \in \left\{0, 1, \dots, \binom{n}{w}-1\right\}$ is a sequence which designates the indices of all the codewords of the extended code $\mathcal{C}_w$ with Hamming weight w; this sequence is expressed in an arbitrary order, and $\hat{l}_i \in (l_0, l_1, \dots, l_{i-1})$ for $1 \leq i \leq \binom{n}{w}-1$. In order to obtain the tightest upper bound on the LHS of (42) in this approach, one has to order the error events such that the correlation coefficients which correspond to codewords $\mathbf{c}_{l_i}$ and $\mathbf{c}_{\hat{l}_i}$ get their maximal available value, which is $1 - \frac{n}{w(n-w)}$ (see [16, Appendix D]). Let us designate this value by $\rho_{w,w}$, i.e.,

$$\rho_{w,w} = 1 - \frac{n}{w(n-w)} \quad , \ w \notin \{0, n\}. \tag{43}$$

Hence, based on the geometry of Fig. 2, if $z_1 \le \sqrt{nE_{\mathrm{s}}}$, we can rewrite (42) as

$$\begin{aligned}
&\Pr\left(\bigcup_{j:\, w_{\mathrm{H}}(\mathbf{c}_j^w)=w} E_{0,j}(z_1),\ \mathbf{y}\in C_n(\theta)\mid z_1\right)\\
&\le \Pr\left(\beta_w(z_1)\le Z_2\le r_{z_1}, V\le r_{z_1}^2-Z_2^2\mid z_1\right)\\
&\quad+\left[\binom{n}{w}-1\right]\cdot\Pr\Bigg(\beta_w(z_1)\le Z_2\le r_{z_1},\\
&\qquad\qquad -r_{z_1}\le Z_3\le \min\Big\{l_{w,w}(z_1,Z_2), r_{z_1}\Big\},\\
&\qquad\qquad W\le r_{z_1}^2-Z_2^2-Z_3^2\mid z_1\Bigg)
\end{aligned}\tag{44}$$

where

$$l_{w,w}(z_1,z_2)=\frac{\beta_w(z_1)-\rho_{w,w}z_2}{\sqrt{1-\rho_{w,w}^2}}.\tag{45}$$

By replacing the first term on the RHS of (40) with the RHS of (44), plugging the result in (7) and averaging w.r.t. Z_1 finally gives the following upper bound on the block error probability:

$$\begin{aligned}
&\Pr(E)\\
&\le \min_w\Bigg\{\Pr\Big(Z_1\le\sqrt{nE_{\mathrm{s}}},\ \beta_w(Z_1)\le Z_2\le r_{Z_1},\\
&\qquad V\le r_{Z_1}^2-Z_2^2\Big)\\
&\qquad+\binom{n}{w}\Pr\Big(Z_1\le\sqrt{nE_{\mathrm{s}}},\ \beta_w(Z_1)\le Z_2\le r_{Z_1},\\
&\qquad\qquad -r_{Z_1}\le Z_3\le\min\Big\{l_{w,w}(Z_1,Z_2), r_{z_1}\Big\},\\
&\qquad\qquad W\le r_{Z_1}^2-Z_2^2-Z_3^2\Big)\\
&\qquad+\sum_{h\neq w}A_h\Pr\Bigg(Z_1\le\sqrt{nE_{\mathrm{s}}},\ \beta_h(Z_1)\le Z_2\le r_{Z_1},\\
&\qquad\qquad -r_{Z_1}\le Z_3\le\min\Big\{l_{w,h}(Z_1,Z_2), r_{z_1}\Big\},\\
&\qquad\qquad W\le r_{Z_1}^2-Z_2^2-Z_3^2\Bigg)\Bigg\}\\
&\quad+\Pr\left(z_1\le\sqrt{nE_{\mathrm{s}}},\ Y\ge r_{Z_1}^2\right)+\Pr\left(Z_1>\sqrt{nE_{\mathrm{s}}}\right).
\end{aligned}\tag{46}$$

Rewriting the RHS of (46) in terms of probability density functions, the AHP bound gets the form

$$\begin{aligned}
\Pr(E)\le\min_w\Bigg\{&\int_{-\infty}^{\sqrt{nE_{\mathrm{s}}}}\Bigg[\int_{\beta_w(z_1)}^{r_{z_1}}f_{Z_2}(z_2)\int_0^{r_{z_1}^2-z_2^2}f_V(v)\,dv\cdot dz_2\\
&+\binom{n}{w}\int_{\beta_w(z_1)}^{r_{z_1}}\int_{-r_{z_1}}^{\min\{l_{w,w}(z_1,z_2),r_{z_1}\}}f_{Z_2,Z_3}(z_2,z_3)\\
&\qquad\int_0^{r_{z_1}^2-z_2^2-z_3^2}f_W(w)\,dw\cdot dz_2\cdot dz_3\\
&+\sum_{\substack{h:\,\beta_h(z_1)<r_{z_1}\\ h\neq w}}\Bigg(A_h\int_{\beta_h(z_1)}^{r_{z_1}}\int_{-r_{z_1}}^{\min\{l_{w,h}(z_1,z_2),r_{z_1}\}}
\end{aligned}$$

$$f_{Z_2,Z_3}(z_2,z_3)\int_0^{r_{z_1}^2-z_2^2-z_3^2} f_W(w)\,dw\,dz_2\,dz_3\Bigg)$$
$$+1-\gamma\left(\frac{n-1}{2},\frac{r_{z_1}^2}{2\sigma^2}\right)\Bigg]f_{Z_1}(z_1)dz_1\Bigg\}$$
$$+Q\left(\sqrt{\frac{2nRE_{\rm b}}{N_0}}\right) \tag{47}$$

where V and W are introduced at the end of Section II-C (after Eq. (37)), and the last term in (47) follows from (13).

III. The Error Exponents of the Improved Versions of the Tangential Sphere Bound

The ITSB and the AHP bound are derived in [16], [17] as upper bounds on the ML decoding error probability of binary linear block codes which are transmitted over the binary-input AWGN channel. In the following, we discuss the exponential behavior of the new upper bounds for arbitrary ensembles of binary linear block codes, and compare it to the error exponent of the TSB. The following lemma is also noted in [17]:

Lemma 3.1: Let $\mathcal{C}$ be a binary linear block code, and let us denote by ITSB($\mathcal{C}$) and TSB($\mathcal{C}$) the ITSB and TSB, respectively, on the decoding error probability of $\mathcal{C}$. Then

$$\text{ITSB}(\mathcal{C}) \le \text{TSB}(\mathcal{C}).$$

Proof: Since $\Pr(A,B) \le \Pr(A)$ for arbitrary events A and B, the lemma follows immediately by comparing the bounds on the RHS of (10) and (25) which correspond to the TSB and the ITSB, respectively. ■

Corollary 3.1: The ITSB can not exceed the value of the TSB while referring to the average error probability of an ensemble of binary linear block codes under ML decoding.

Lemma 3.2: The AHP bound cannot be looser than the TSB in the limit where we let the block length tend to infinity.

Proof: To show this, we rely on (46) and extend the code by referring to the layer of Hamming weight $w = n$. Hence, the extended code contains at most one additional codeword whose Hamming weight is n (as compared to the codebook of the original code); this possible extension of the codebook by a single codeword has no impact on the exponential behavior of the decoding error probability for long enough codes. The resulting upper bound is evidently not tighter than the AHP bound (which carries an optimization over w), and on the other hand, it is at least as tight as the TSB (since the joint probability of two events cannot exceed the probabilities of these individual events). ■

The extension of Lemma 3.2 to ensembles of codes is straightforward (by calculating the statistical expectation over the codes of an ensemble, Lemma 3.2 also holds for ensembles). Hence, it follows that the error exponents of both the AHP bound and the ITSB are larger or equal to the error exponent of the TSB. In the following, we introduce a lower bound on both the ITSB and the AHP bound. It serves as an intermediate stage to get our main result.

Lemma 3.3: Let $\mathcal{C}$ designate an ensemble of binary linear block codes of length n, whose transmission takes place over an AWGN channel. Let A_h be the number of codewords of Hamming weight h (where $0 \le h \le n$), and let $\mathbb{E}_{\mathcal{C}}$ designate the statistical expectation over the codebooks of an ensemble $\mathcal{C}$. Then both the ITSB and AHP upper bounds on the average ML decoding error probability of $\mathcal{C}$ are lower bounded by

$$\psi(\mathcal{C}) \triangleq \min_w \Bigg\{ \Pr\Big(Z_1 \le \sqrt{nE_{\rm s}},\ \beta_w(Z_1) \le Z_2 \le r_{Z_1},$$
$$V \le r_{Z_1}^2 - Z_2^2 \Big)$$
$$+ \sum_h \Bigg\{ \mathbb{E}_{\mathcal{C}}[A_h] \Pr\Big(Z_1 \le \sqrt{nE_{\rm s}},\ \beta_h(Z_1) \le Z_2 \le r_{Z_1},$$
$$-r_{Z_1} \le Z_3 \le \min\Big\{ l_{w,h}(Z_1,Z_2), r_{Z_1} \Big\},$$
$$W \le r_{Z_1}^2 - Z_2^2 - Z_3^2 \Big) \Bigg\}$$
$$+ \Pr\left(Z_1 \le \sqrt{nE_{\rm s}},\ Y \ge r_{Z_1}^2 \right) \Bigg\} \tag{48}$$

where $l_{w,h}(z_1, z_2)$ is defined in (41).

Proof: By comparing (46) with (48), it is easily verified that the RHS of (48) is not larger than the RHS of (46) (actually, the RHS of (48) is just the AHP *without any extension of the code*). Referring to the ITSB, we get by calculating the statistical expectation of the RHS of (35) w.r.t. the codes of the ensemble $\mathcal{C}$

$$
\begin{aligned}
&\mathrm{ITSB}(\mathcal{C}) \\
&\overset{(a)}{=} \Pr\left(Z_1 \le \sqrt{nE_{\mathrm{s}}},\, \beta_{\min}(Z_1) \le Z_2 \le r_{Z_1},\, V \le r_{Z_1}^2 - Z_2^2 \right) \\
&\quad + \sum_h \Bigg\{ \mathbb{E}_{\mathcal{C}}[A_h] \Pr\Big(Z_1 \le \sqrt{nE_{\mathrm{s}}},\, \beta_h(Z_1) \le Z_2 \le r_{Z_1}, \\
&\qquad\qquad -r_{Z_1} \le Z_3 \le \min\Big\{ l_h(Z_1, Z_2), r_{Z_1} \Big\}, \\
&\qquad\qquad W \le r_{Z_1}^2 - Z_2^2 - Z_3^2 \Big) \Bigg\} \\
&\quad + \Pr\left(Z_1 \le \sqrt{nE_{\mathrm{s}}},\, Y \ge r_{Z_1}^2 \right) + \Pr\left(Z_1 > \sqrt{nE_{\mathrm{s}}} \right) \\
&\overset{(b)}{\ge} \min_w \Bigg\{ \Pr\left(Z_1 \le \sqrt{nE_{\mathrm{s}}},\, \beta_w(Z_1) \le Z_2 \le r_{Z_1}, \right. \\
&\qquad\qquad \left. V \le r_{Z_1}^2 - Z_2^2 \right) \\
&\qquad + \sum_h \Bigg\{ \mathbb{E}_{\mathcal{C}}[A_h] \Pr\Big(Z_1 \le \sqrt{nE_{\mathrm{s}}},\, \beta_h(Z_1) \le Z_2 \le r_{Z_1}, \\
&\qquad\qquad -r_{Z_1} \le Z_3 \le \min\Big\{ l_{w,h}(Z_1, Z_2), r_{Z_1} \Big\}, \\
&\qquad\qquad W \le r_{Z_1}^2 - Z_2^2 - Z_3^2 \Big) \Bigg\} \\
&\qquad + \Pr\left(Z_1 \le \sqrt{nE_{\mathrm{s}}},\, Y \ge r_{Z_1}^2 \right) \Bigg\} + \Pr\left(Z_1 > \sqrt{nE_{\mathrm{s}}} \right) \\
&\overset{(c)}{>} \psi(\mathcal{C}).
\end{aligned}
\tag{49}
$$

Equality (a) in (49) follows from calculating the statistical expectation of the RHS of (35) w.r.t. the codes from the ensemble $\mathcal{C}$, and also from the linearity of this expression w.r.t. the distance spectrum. Inequality (b) follows by replacing $w = d_{\min}$ with a minimization w.r.t. w, and since the ITSB is a monotonically decreasing function w.r.t. the correlation coefficients (see Appendix III where we rely here on (33), (34), (39) and (43) to get that $\rho_h \le \rho_{d_{\min},h}$ for all values of h). Inequality (c) follows by removing the last positive term, $\Pr\left(Z_1 > \sqrt{nE_{\mathrm{s}}}\right)$, which gives (48). ■

In [16] and [17], the RHS of (46) and (35), respectively, were evaluated by integrals, which results in the upper bounds (47) and (37). In [1, Section D], Divsalar introduced an alternative way to obtain a simple, yet an asymptotically identical version of the TSB by using the Chernoff bounding technique. Using this technique, we obtain the exponential version of $\psi(\mathcal{C})$ which happens to coincide with that one of the TSB (see Appendix I). In the following, we state the main result of this paper.

Theorem 3.1: (*The error exponents of the AHP bound and the ITSB coincide with the error exponent of the TSB*) The TSB, ITSB and the AHP bound possess the same error exponent, which is given by

$$
E(c) = \min_{0<\delta\le 1} \left\{ \frac{1}{2} \ln\left(1 - \gamma + \gamma e^{-2r(\delta)} \right) + \frac{\gamma \Delta^2 c}{1 + \gamma \Delta^2} \right\} \tag{50}
$$

where

$$\gamma = \gamma(\delta) \triangleq \frac{1-\delta}{\delta}\left[\sqrt{\frac{c}{c_0(\delta)} + (1+c)^2 - 1} - (1+c)\right] \tag{51}$$

$$c_0(\delta) \triangleq \left(1 - e^{-2r(\delta)}\right)\frac{1-\delta}{2\delta}. \tag{52}$$

Here, δ designates the normalized Hamming weight

$$c \triangleq \frac{E_{\mathrm{s}}}{N_0}, \quad \Delta \triangleq \sqrt{\frac{\delta}{1-\delta}}$$

and $r(\delta)$ denotes the asymptotic growth rate of the (average) distance spectrum of the code (ensemble) in terms of δ.

Proof: The exponential version of $\psi(\mathcal{C})$ in (48) is identical to the exponential version of the TSB (see Appendices I and II). Since $\psi(\mathcal{C})$ does not exceed the AHP bound and the ITSB (see Lemma 3.3), this implies that the error exponents of the AHP and the ITSB are not larger than the error exponent of the TSB. On the other hand, from Lemmas 3.1 and 3.2 it follows that asymptotically, both the AHP and the ITSB are at least as tight as the TSB, so their error exponents are at least as large as the error exponent of the TSB. Combining these results we obtain that the error exponents of the ITSB, AHP and the TSB are all identical. In [1], Divsalar shows that the error exponent of the TSB is determined by (50)–(52), which concludes the proof of the theorem. ■

Remark 3.1: The upper bound on the bit error probability in [11] is exactly the same as the TSB on the block error probability by Poltyrev [9], except that the average distance spectrum $\{A_h\}$ of the ensemble is now replaced by the sequence $\{A'_h\}$ where

$$A'_h = \sum_{w=0}^{nR}\left(\frac{w}{nR}\right)A_{w,h}\ , \quad h \in \{0,\ldots,n\}$$

and $A_{w,h}$ denotes the average number of codewords encoded by information bits of Hamming weight w and having a Hamming weight (after encoding) which is equal to h. Since

$$A_h = \sum_{w=0}^{nR} A_{w,h}$$

then

$$\frac{A_h}{nR} \le A'_h \le A_h\ , \quad h \in \{0,\ldots,n\}.$$

The last inequality therefore implies that the replacement of the distance spectrum $\{A_h\}$ by $\{A'_h\}$ (for the analysis of the bit error probability) does not affect the asymptotic growth rate of $r(\delta)$ where $\delta \triangleq \frac{h}{n}$, and hence, the error exponents of the TSB on the block and bit error probabilities coincide.

Remark 3.2: In [19], Zangl and Herzog suggest a modification of the TSB on the bit error probability. Their basic idea is to tighten the upper bound on the bit error probability when the received vector $\mathbf{y}$ falls outside the cone $\mathcal{R}$ on the RHS of (3) (see Fig. 1). In the derivation of the version of the TSB on the bit error probability, as suggested by Sason and Shamai [11], the conditional bit error probability in this case was upper bounded by 1, where Zangl and Herzog [19] refine the bound and provide a tighter bound on the conditional bit error probability when the vector $\mathbf{y}$ falls in the bad region (i.e., when it is outside the cone in Fig. 1). Though this modification tightens the bound on the bit error probability at low SNR (as exemplified in [19] for some short linear block codes), it has no effect on the error exponent. The reason is simply because the conditional bit error probability in this case cannot be below $\frac{1}{nR}$ (i.e., one over the dimension of the code), so the bound should still possess the same error exponent. This shows that the TSB versions on the bit error probability, as suggested in [11] and [19], have the same error exponents as of the TSB.

Corollary 3.2: The error exponents of the TSB on the bit error probability coincides with the error exponent of the TSB on the block error probability. Moreover, the error exponents of the TSB on the bit error probability, as suggested by Sason and Shamai [11] and refined by Zangl and Herzog [19], coincide. The common value of these error exponents is explicitly given in Theorem 3.1.

IV. Summary and Conclusions

The tangential sphere bound (TSB) of Poltyrev [9] and its recent improvements by Yousefi et al. ([16], [17]) serve as useful upper bounds on the ML decoding error probability of binary linear block codes, transmitted over a binary-input AWGN channel. However, in the random coding setting, the TSB fails to reproduce the random coding error exponent [6]; the larger the code rate is, the more significant becomes the gap between the error exponent of the TSB and the random coding error exponent of Gallager [6] (see Fig. 3, and the plots in [9, Figs. 2–4]). In this respect, we note that the expression for the error exponent of the TSB, as derived by Divsalar [1], is significantly easier for numerical calculations than the original expression of this error exponent which was provided by Poltyrev [9, Theorem 2]. The analysis made by Divsalar is also more general in the sense that it applies to an arbitrary ensemble, so it is not only restricted to the ensemble of fully random block codes.

The recently introduced bounds by Yousefi et al. ([16]–[18]) solely depend on the distance spectrum of the code (or on their input-output weight enumerators for the analysis of the bit error probability). Though these new bounds were previously exemplified (see [16]–[18]) to slightly tighten the TSB for short binary linear block codes, their error exponents were not considered yet. The focus of this paper is on the analysis of the error exponents of these new bounds for an arbitrary ensemble of binary linear block codes; it is given in terms of the asymptotic growth rate of the average distance spectrum for the considered ensemble.

Putting the results reported by Divsalar [1] with the main result in this paper (see Theorem 3.1), we conclude that the error exponents of the simple bound of Divsalar [1], the first version of Duman and Salehi bounds [2], and the Chernoff versions of the TSB [9] and its recent improvements by Yousefi et al. [16]–[18] all coincide. This conclusion holds for an arbitrary ensemble of binary linear block codes (e.g., turbo codes, LDPC codes etc.) where we let the block length tend to infinity, in addition to the ensemble of fully random block codes (whose distance spectrum is binomially distributed). Moreover, the TSB versions for the bit error probability, as provided in [11] and [19], have error exponents which coincide; their common value is equal to the error exponent of the TSB for the block error probability. Based on Theorem 3.1, it follows that for any value of SNR, the same value of the normalized Hamming weight dominates the exponential behavior of the TSB and its two improved versions. In the asymptotic case where we let the block length tend to infinity, the dominating normalized Hamming weight can be explicitly calculated in terms of the SNR; this calculation is based on finding the value of the normalized Hamming weight δ which achieves the minimal value of the RHS of (50), where this value clearly depends on the asymptotic growth rate of the distance spectrum of the ensemble under consideration. A similar calculation of this critical weight as a function of the SNR was done in [4] while referring to the ensemble of fully random block codes and the union bound.

In a companion paper [14], new upper bounds on the block and bit error probabilities of linear block codes are derived. These bounds improve the tightness of the Shulman and Feder bound [13] and therefore also reproduce the random coding error exponent.

Acknowledgment

The authors are grateful to the three anonymous reviewers for their constructive comments.

Appendix I

The exponential behavior of $\psi(\mathcal{C})$ in (48)

In the following, the exponential behavior of the RHS of (48) is obtained by using the Chernoff bound for $\psi(\mathcal{C})$.

Note that the geometrical region of the TSB corresponds to a double sided circular cone. For the derivation of the bound for the circular cone in Fig. 1, we refer to the event where $Z_1 \leq \sqrt{nE_{\mathrm{s}}}$; however, since $Z_1 \sim N(0, \frac{N_0}{2})$, then this boundary effect does not have any implication on the exponential behavior of the function $\psi(\mathcal{C})$ for large values of n (as also noted in [1, p. 23]). To simplify the analysis, we therefore do not take into consideration of this boundary effect for large values of n. Let $\widetilde{\psi}(\mathcal{C})$ designate the function which is obtained by removing the event $Z_1 \leq \sqrt{nE_{\mathrm{s}}}$ from the expression for $\psi(\mathcal{C})$ (see the RHS of (48)); then, the exponential behavior of $\psi(\mathcal{C})$ and $\widetilde{\psi}(\mathcal{C})$ asymptotically coincide for large values of n.

Let us designate the normalized Gaussian noise vector by $\underline{\Lambda} = (\Lambda_1, \ldots, \Lambda_n)$, i.e., $(\Lambda_1, \ldots, \Lambda_n) = \sqrt{\frac{2}{N_0}}(Z_1, \ldots, Z_n)$, and define $\eta \triangleq \tan^2 \theta$. The Gaussian random vector has n orthogonal components which are therefore statistically

independent. Without abuse of notation, let $\lambda_1, \ldots, \lambda_n$ designate in this appendix realizations of the random variables $\Lambda_1, \ldots, \Lambda_n$, respectively (note that the sequence $\{\lambda_i\}$ was used in a different context for the derivation of the ITSB, where the λ_i's there are integer numbers). From (4) and (41), and by multiplying all signals by $\sqrt{\frac{2}{N_0}}$ (which follows from the scaling above), the following holds for BPSK modulated signals:

$$
\begin{aligned}
& r = \sqrt{2nc\eta}, \quad r_{\lambda_1} = \sqrt{\eta}\left(\sqrt{2nc} - \lambda_1\right) \\
& \beta_h(\lambda_1) = \left(\sqrt{2nc} - \lambda_1\right)\sqrt{\frac{h}{n-h}} \\
& l_{w,h}(\lambda_1, \lambda_2) = \frac{\beta_w(\lambda_1) - \rho_{w,h}\,\lambda_2}{\sqrt{1-\rho_{w,h}^2}} \\
& c \triangleq \frac{E_{\mathrm{s}}}{N_0}.
\end{aligned}
\tag{I.1}
$$

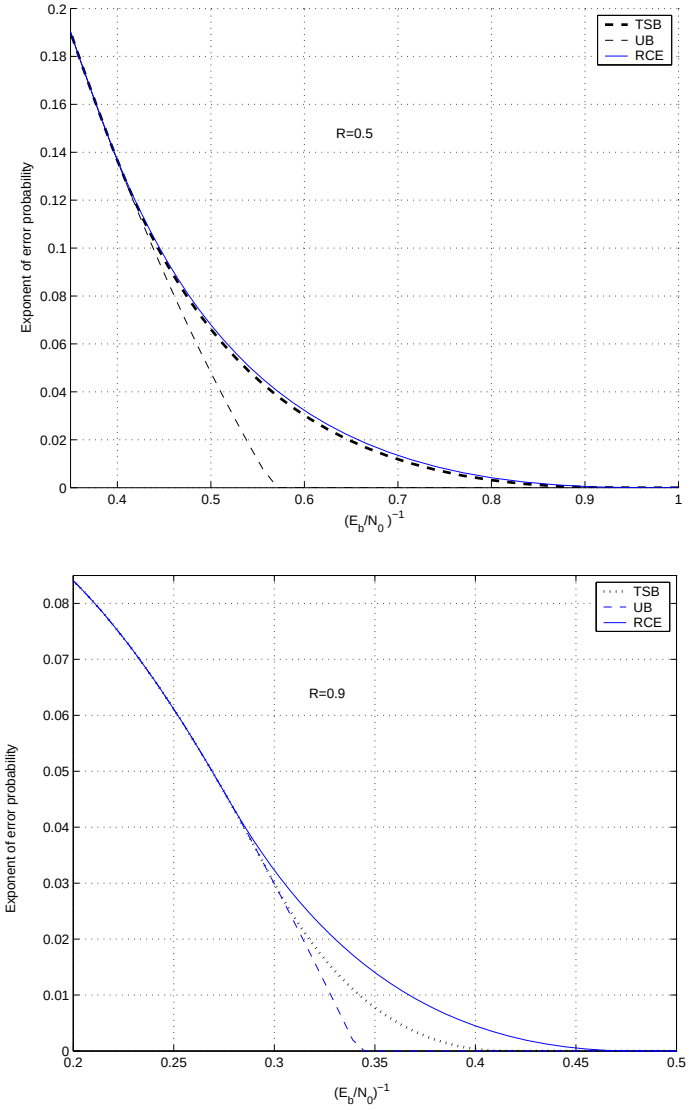

Fig. 3. Comparison between the error exponents for random block codes which are based on the union bound (UB), the tangential sphere bound (TSB) of Poltyrev [9] (which according to Theorem 3.1 is identical to the error exponents of the ITSB and AHP bounds), and the random coding bound (RCE) of Gallager [6]. The upper and lower plots refer to code rates of 0.5 and 0.9 bits per channel use, respectively. The error exponents are plotted versus the reciprocal of the energy per bit to the one-sided spectral noise density.

Hence, we obtain from (48) and the above discussion

$$\begin{aligned}\widetilde{\psi}(\mathcal{C}) = &\min_{w}\Bigg\{\Pr\left(\sum_{i=2}^{n}\Lambda_i^2 \le r_{\Lambda_1}^2, \Lambda_2 \ge \beta_w(\Lambda_1)\right)\\ &+\sum_{h=1}^{n} A_h \Pr\left(\sum_{i=2}^{n}\Lambda_i^2 \le r_{\Lambda_1}^2,\, \Lambda_2 \ge \beta_h(\Lambda_1),\, \Lambda_3 \ge -l_{w,h}(\Lambda_1,\Lambda_2)\right)\\ &+\Pr\left(\sum_{i=2}^{n}\Lambda_i^2 \ge r_{\Lambda_1}^2\right)\Bigg\}. \end{aligned} \tag{I.2}$$

At this point, we upper bound the RHS of (I.2) by the Chernoff bounds, namely, for three random variables V, W and Z

$$\begin{aligned}&\Pr\left(V \ge 0\right) \le \mathbb{E}\left[e^{pV}\right], \quad p \ge 0\\ &\Pr\left(W \le 0,\, V \ge 0\right) \le \mathbb{E}\left[e^{qW+uV}\right], \;\; q \le 0,\; u \ge 0\\ &\Pr\left(W \le 0,\, V \ge 0,\, Z \ge 0\right) \le \mathbb{E}\left[e^{tW+sV+kZ}\right],\\ &\qquad t \le 0,\; s \ge 0,\; k \ge 0. \end{aligned} \tag{I.3}$$

The Chernoff versions of the first and last terms on the RHS of (I.2) are introduced in [1, Eqs.(134)–(137)], and are given by

$$\Pr\left(\sum_{i=2}^{n}\Lambda_i^2 \ge r_{\Lambda_1}^2\right) \le \sqrt{\frac{1-2p}{1+2p\eta}}\, e^{-nE_1(c,p,\eta)},\; p \ge 0 \tag{I.4}$$

$$\begin{aligned}&\Pr\left(\sum_{i=2}^{n}\Lambda_i^2 \le r_{\Lambda_1}^2,\, \Lambda_2 \ge \beta_w(\Lambda_1)\right)\\ &\quad\le \sqrt{\frac{1-2q}{1+2q\eta}}\, e^{-nE_2(c,q,\frac{w}{n},\eta)},\; -\frac{1}{2\eta} \le q \le 0 \end{aligned} \tag{I.5}$$

where

$$E_1(c,p,\eta) \triangleq \frac{2p\eta c}{1+2p\eta} + \frac{1}{2}\ln(1-2p) \tag{I.6}$$

and

$$\begin{aligned}&E_2(c,q,\delta,\eta)\\ &\quad\triangleq\quad c\left(\frac{2q\eta + (1-2q)\sqrt{\frac{\delta}{1-\delta}}}{1+2q\eta+(1-2q)\sqrt{\frac{\delta}{1-\delta}}}\right) + \frac{1}{2}\ln(1-2q). \end{aligned} \tag{I.7}$$

Next, by invoking the Chernoff bound (I.3), we get an exponential upper bound on the second term on the RHS of (48). Using the notation

$$\zeta_{w,h} \triangleq \sqrt{\frac{w(n-h)}{h(n-w)}} \tag{I.8}$$

we get (see Appendix II for details)

$$\begin{aligned}&A_h \Pr\left(\sum_{i=2}^{n}\Lambda_i^2 \le r_{\Lambda_1}^2,\, \Lambda_2 \ge \beta_h(\Lambda_1),\, \Lambda_3 \ge -l_{w,h}(\Lambda_1,\Lambda_2)\right)\\ &\le \sqrt{\frac{1-2t}{1+2t\eta}}\, e^{-g(c,t,k,s,\eta,h,n)},\\ &\qquad -\frac{1}{2\eta} \le t \le 0,\; k \ge 0,\; s \ge 0 \end{aligned} \tag{I.9}$$

where

$$\begin{aligned} &g(c,t,k,s,\eta,h,n) \\ &\triangleq \frac{4t\eta nc + 2\sqrt{2nc}\left(s - \frac{k\zeta_{w,h}}{\sqrt{1-\rho_{w,h}^2}}\right)\Delta_h - \Delta_h^2\left(s - \frac{k\zeta_{w,h}}{\sqrt{1-\rho_{w,h}^2}}\right)^2}{2(1+2t\eta)} \\ &\quad - \frac{\left(s - \frac{k\rho_{w,h}}{\sqrt{1-\rho_{w,h}^2}}\right)^2 + k^2}{2(1-2t)} + \frac{n}{2}\ln(1-2t) - n\, r_n\Big(\frac{h}{n}\Big) \end{aligned} \tag{I.10}$$

and

$$\Delta_h \triangleq \sqrt{\frac{h}{n-h}}, \quad r_n\Big(\frac{h}{n}\Big) \triangleq \frac{\ln A_h}{n}.$$

The next step is to find optimal values for k and s in order to maximize the function g. If $k^* = 0$ then the exponent of $\psi(\mathcal{C})$ is identical to that of the TSB. In order to find the optimal $k \geq 0$ and $s \geq 0$ which maximize g, we consider the aforementioned probabilities by discussing separately the three cases where $h < w,\ h > w$ and $h = w$.

Case 1: $h = w$. In this case, it follows from (I.8) that $\zeta_{w,h} = \zeta_{w,w} = 1$, and we get

$$\begin{aligned} &A_w \Pr\left(\sum_{i=2}^{n}\Lambda_i^2 \leq r_{\Lambda_1}^2,\ \Lambda_2 \geq \beta_w(\Lambda_1),\ \Lambda_3 \geq -l_{w,w}(\Lambda_1,\Lambda_2)\right) \\ &\leq \sqrt{\frac{1-2t}{1+2t\eta}}\, e^{-g(c,t,k,s,\eta,w,n)}, \\ &\qquad -\frac{1}{2\eta} \leq t \leq 0,\ k \geq 0,\ s \geq 0 \end{aligned} \tag{I.11}$$

where

$$\begin{aligned} &g(c,t,k,s,\eta,w,n) \\ &= \frac{4t\eta nc + 2\sqrt{2nc}\left(s - \frac{k}{\sqrt{1-\rho_{w,w}^2}}\right)\Delta_w - \Delta_w^2\left(s - \frac{k}{\sqrt{1-\rho_{w,w}^2}}\right)^2}{2(1+2t\eta)} \\ &\quad - \frac{\left(s - \frac{k\rho_{w,w}}{\sqrt{1-\rho_{w,w}^2}}\right)^2}{2(1-2t)} - \frac{k^2}{2(1-2t)} + \frac{n}{2}\ln(1-2t) - \ln(A_w). \end{aligned} \tag{I.12}$$

Let us introduce the parameters

$$\xi = s - \frac{k}{\sqrt{1-\rho_{w,w}^2}} \tag{I.13}$$

$$\tau = s - \frac{k\rho_{w,w}}{\sqrt{1-\rho_{w,w}^2}}. \tag{I.14}$$

From (I.13) and (I.14), we get

$$k = -(\xi - \tau)\alpha \tag{I.15}$$

where

$$\alpha \triangleq \sqrt{\frac{1+\rho_{w,w}}{1-\rho_{w,w}}}. \tag{I.16}$$

Hence, the Chernoff bounding technique gives

$$\begin{aligned} &\Pr\left(\sum_{i=2}^{n}\Lambda_i^2 \leq r_{\Lambda_1}^2,\ \Lambda_2 \geq \beta_w(\Lambda_1),\ \Lambda_3 \geq -l_{w,w}(\Lambda_1,\Lambda_2)\right) \\ &\leq \sqrt{\frac{1-2t}{1+2t\eta}} e^{-g_1(c,t,\xi,\tau,\eta,w,n)}, \quad -\frac{1}{2\eta} \leq t \leq 0 \end{aligned} \tag{I.17}$$

where

$$g_1(c,t,\xi,\tau,\eta,h,n) = \frac{4t\eta nc + 2\sqrt{2nc}\xi\Delta_w - \Delta_w^2\xi^2}{2(1+2t\eta)} - \frac{\tau^2}{2(1-2t)} - \frac{(\xi-\tau)^2\alpha^2}{2(1-2t)} + \frac{n}{2}\ln(1-2t). \quad \text{(I.18)}$$

Maximizing the RHS of (I.17) w.r.t. τ yields

$$\begin{aligned} \frac{\partial g_1}{\partial \tau} &= -\frac{\tau}{1-2t} + \frac{(\xi-\tau)\alpha^2}{1-2t} = 0 \\ \Rightarrow \tau^* &= \frac{\alpha^2\xi^*}{1+\alpha^2}. \end{aligned} \quad \text{(I.19)}$$

Notice that $\frac{\partial^2 g_1}{\partial \tau^2} < 0$, hence plugging τ^* in (I.18) maximizes g_1. Substituting τ^* into (I.18) gives

$$\begin{aligned} g_2(c,t,\xi,\eta,w,n) \triangleq & g_1(c,t,\xi,\tau^*,\eta,w,n) \\ = & \frac{4t\eta nc + 2\sqrt{2nc}\Delta_w\xi - \Delta_w^2\xi^2}{2(1+2t\eta)} \\ & - \frac{\frac{\alpha^2}{1+\alpha^2}\xi^2}{2(1-2t)} + \frac{n}{2}\ln(1-2t). \end{aligned}$$

A differentiation of g_2 w.r.t. ξ and an introduction of the new parameter $\epsilon \triangleq \frac{\alpha^2}{1+\alpha^2}$ gives

$$\begin{aligned} \frac{\partial g_2}{\partial \xi} &= \frac{\sqrt{2nc}\Delta_w - \Delta_w^2\xi}{1+2t\eta} - \frac{\epsilon\xi}{1-2t} = 0 \\ \xi^* &= \frac{\sqrt{2nc}\Delta_w(1-2t)}{\Delta_w^2(1-2t) + \epsilon(1+2t\eta)}. \end{aligned}$$

Again, $\frac{\partial^2 g_2}{\partial \xi^2} < 0$, so $\xi = \xi^*$ maximizes the function g_2. From (I.19), $\xi^* - \tau^* > 0$. Since α is non-negative, we get that k^* in (I.15) is not-positive. But since from (I.9), $k \geq 0$, this yields that the optimal value of k is equal to zero. From the Chernoff bound in (I.3), an optimality of k when it is set to zero implies that asymptotically, as $n \to \infty$

$$\begin{aligned} & \Pr\left(\sum_{i=2}^{n}\Lambda_i^2 \leq r_{\Lambda_1}^2,\, \Lambda_2 \geq \beta_w(\Lambda_1),\, \Lambda_3 \geq -l_{w,w}(\Lambda_1,\Lambda_2)\right) \\ & \doteq \Pr\left(\sum_{i=2}^{n}\Lambda_i^2 \leq r_{\Lambda_1}^2,\, \Lambda_2 \geq \beta_w(\Lambda_1)\right). \end{aligned} \quad \text{(I.20)}$$

Case 2: $h > w$. In this case, it follows from (39) that $\rho_{w,h} = \sqrt{\frac{w(n-h)}{h(n-w)}}$. Hence, for this case, we get from (I.8) that $\rho_{w,h} = \zeta_{w,h}$. Replacing $\rho_{w,h}$ by $\zeta_{w,h}$ in (I.10) gives

$$\begin{aligned} & g(c,t,k,s,\eta,h,n) \\ & = \frac{4t\eta nc + 2\sqrt{2nc}\left(s - \frac{k\zeta_{w,h}}{\sqrt{1-\zeta_{w,h}^2}}\right)\Delta_h - \Delta_h^2\left(s - \frac{k\zeta_{w,h}}{\sqrt{1-\zeta_{w,h}^2}}\right)^2}{2(1+2t\eta)} \\ & \quad - \frac{\left(s - \frac{k\zeta_{w,h}}{\sqrt{1-\zeta_{w,h}^2}}\right)^2}{2(1-2t)} - \frac{k^2}{2(1-2t)} + \frac{n}{2}\ln(1-2t) - nr_n\Big(\frac{h}{n}\Big). \end{aligned}$$

In the following, we introduce the parameters:

$$\xi \triangleq s - \frac{k\zeta_{w,h}}{\sqrt{1-\zeta_{w,h}^2}}, \qquad \tau \triangleq k.$$

Optimization over τ yields $\tau^* = 0$, so $k^* = 0$, and asymptotically (as we let n tend to infinity), one gets the following equality in terms of the exponential behaviors:

$$\begin{aligned}&\Pr\left(\sum_{i=2}^{n}\Lambda_i^2 \le r_{\Lambda_1}^2,\ \Lambda_2 \ge \beta_h(\Lambda_1),\ \Lambda_3 \ge -l_{w,h}(\Lambda_1,\Lambda_2)\right)\\ &\doteq \Pr\left(\sum_{i=2}^{n}\Lambda_i^2 \le r_{\Lambda_1}^2,\ \Lambda_2 \ge \beta_h(\Lambda_1)\right).\end{aligned} \tag{I.21}$$

Case 3: $h < w$. From (39), it follows that for these values of h, $\rho_{w,h} = \sqrt{\frac{h(n-w)}{w(n-h)}}$, so we get from (I.8) that $\rho_{w,h} < \zeta_{w,h}$. Define

$$\xi \triangleq s - \frac{k\zeta_{w,h}}{\sqrt{1-\rho_{w,h}^2}}, \qquad \tau \triangleq s - \frac{k\rho_{w,h}}{\sqrt{1-\rho_{w,h}^2}}.$$

This gives $k = -(\xi - \tau)\alpha'$ where

$$\alpha' \triangleq \frac{\sqrt{1-\rho_{w,h}^2}}{\zeta_{w,h} - \rho_{w,h}}.$$

Since in this case $\rho_{w,h} < \zeta_{w,h}$, then $\alpha' > 0$. Similarly to the arguments in case 1, we get again that the optimal value for k is $k^* = 0$ which gives again (I.21) (i.e., in the limit where the block length tends to infinity, the exponential behavior of the LHS and RHS of (I.21) coincide).

APPENDIX II
DERIVATION OF THE CHERNOFF BOUND IN (I.9) WITH THE FUNCTION g IN (I.10)

Using the Chernoff bound (I.3) and defining

$$\Delta_w \triangleq \sqrt{\frac{w}{n-w}} \tag{II.1}$$

we get

$$\begin{aligned}&\Pr\left(\sum_{i=2}^{n}\Lambda_i^2 \le r_{\Lambda_1}^2,\ \Lambda_2 \ge \beta_h(\Lambda_1),\ \Lambda_3 \ge -l_{w,h}(\Lambda_1,\Lambda_2)\right)\\ &\overset{(a)}{\le} \mathbb{E}\left[e^{t\left(\sum_{i=2}^{n}\Lambda_i^2 - r_{\Lambda_1}^2\right) + s(\Lambda_2 - \beta_h(\Lambda_1)) + k(\Lambda_3 + l_{w,h}(\Lambda_1,\Lambda_2))}\right],\\ &\qquad\qquad t \le 0,\ s \ge 0,\ k \ge 0\\ &\overset{(b)}{=} \mathbb{E}\Bigg[e^{t\left(\sum_{i=2}^{n}\Lambda_i^2 - \eta(\sqrt{2nc}-\Lambda_1)^2\right) + s\left(\Lambda_2 - \Delta_h(\sqrt{2nc}-\Lambda_1)\right)}\\ &\qquad\qquad \cdot e^{k\left(\Lambda_3 + \frac{\Delta_w(\sqrt{2nc}-\Lambda_1) - \rho_{w,h}\Lambda_2}{\sqrt{1-\rho_{w,h}^2}}\right)}\Bigg]\\ &= \mathbb{E}\Bigg[e^{t\sum_{i=2}^{n}\Lambda_i^2 - t\eta\Lambda_1^2 - 2tn\eta c + 2\eta t\sqrt{2nc}\Lambda_1 + s\Lambda_2 - s\Delta_h\sqrt{2nc} + s\Delta_h\Lambda_1}\\ &\qquad\qquad \cdot e^{k\Lambda_3 + \frac{k\Delta_w\sqrt{2nc}}{\sqrt{1-\rho_{w,h}^2}} - \frac{k(\Delta_w\Lambda_1 + \rho_{w,h}\Lambda_2)}{\sqrt{1-\rho_{w,h}^2}}}\Bigg]\\ &\overset{(c)}{=} \mathbb{E}\left[e^{t\sum_{i=4}^{n}\Lambda_i^2}\right]\mathbb{E}\left[e^{-t\eta\Lambda_1^2 + \left(2\eta t\sqrt{2nc} + s\Delta_h - \frac{k\Delta_w}{\sqrt{1-\rho_{w,h}^2}}\right)\Lambda_1}\right]\end{aligned}$$

$$\cdot\, \mathbb{E}\left[e^{t\Lambda_2^2+\left(s-\frac{k\rho_{w,h}}{\sqrt{1-\rho_{w,h}^2}}\right)\Lambda_2}\right] \cdot \mathbb{E}\left[e^{t\Lambda_3^2+k\Lambda_3}\right]$$
$$\cdot\, e^{-2tn\eta c-s\Delta_h\sqrt{2nc}+\frac{k\Delta_w\sqrt{2nc}}{\sqrt{1-\rho_{w,h}^2}}}. \tag{II.2}$$

where inequality (a) follows from the Chernoff bound (I.3), equality (b) follows from (I.1), and equality (c) follows from the statistical independence of the components of the normalized noise vector $\underline{\Lambda}$. For a zero-mean and unit-variance Gaussian random variable X, the following equality holds:

$$\mathbb{E}\left[e^{aX^2+bX}\right] = \frac{e^{\frac{b^2}{2(1-2a)}}}{\sqrt{1-2a}}, \quad a \le \frac{1}{2},\ b \in \mathbb{R}. \tag{II.3}$$

Evaluating each term in (II.2) by the equality in (II.3), and substituting $\zeta_{w,h} = \frac{\Delta_w}{\Delta_h}$ which follows from (I.8) and (II.1), gives

$$\mathbb{E}\left[e^{t\sum_{i=4}^{n}\Lambda_i^2}\right] = \left(\frac{1}{\sqrt{1-2t}}\right)^{n-3},\ t \le 0 \tag{II.4}$$

$$\mathbb{E}\left[e^{-t\eta\Lambda_1^2+\left(2\eta t\sqrt{2nc}+s\Delta_h-\frac{k\Delta_w}{\sqrt{1-\rho_{w,h}^2}}\right)\Lambda_1}\right]$$
$$= \frac{1}{\sqrt{1+2t\eta}}\, e^{\frac{\left(2\eta t\sqrt{2nc}+\Delta_h\left(s-\frac{k\zeta_{w,h}}{\sqrt{1-\rho_{w,h}^2}}\right)\right)^2}{2(1+2t\eta)}} \tag{II.5}$$

$$\mathbb{E}\left[e^{t\Lambda_2^2+\left(s-\frac{k\rho_{w,h}}{\sqrt{1-\rho_{w,h}^2}}\right)\Lambda_2}\right] = \frac{1}{\sqrt{1-2t}} \cdot e^{\frac{\left(s-\frac{k\rho_{w,h}}{\sqrt{1-\rho_{w,h}^2}}\right)^2}{2(1-2t)}},$$
$$k \ge 0,\ s \ge 0 \tag{II.6}$$

$$\mathbb{E}\left[e^{t\Lambda_3^2+k\Lambda_3}\right] = \frac{1}{\sqrt{1-2t}}\, e^{\frac{k^2}{2(1-2t)}}, \quad t \le 0,\ k \ge 0. \tag{II.7}$$

From (II.5), straightforward algebra gives

$$\mathbb{E}\left[e^{-t\eta\Lambda_1^2+\left(2\eta t\sqrt{2nc}+s\Delta_h-\frac{k\Delta_w}{\sqrt{1-\rho_{w,h}^2}}\right)\Lambda_1}\right]$$
$$\cdot\, e^{-2tn\eta c-s\Delta_h\sqrt{2nc}+\frac{k\Delta_w\sqrt{2nc}}{\sqrt{1-\rho_{w,h}^2}}}$$
$$= \frac{1}{\sqrt{1+2t\eta}} \cdot \exp\left\{\frac{-4t\eta nc - 2\sqrt{2nc}\left(s-\frac{k\zeta_{w,h}}{\sqrt{1-\rho_{w,h}^2}}\right)\Delta_h}{2(1+2t\eta)}\right\}$$
$$\cdot \exp\left\{\frac{\Delta_h^2\left(s-\frac{k\zeta_{w,h}}{\sqrt{1-\rho_{w,h}^2}}\right)^2}{2(1+2t\eta)}\right\}. \tag{II.8}$$

Plugging (II.4) and (II.6)–(II.8) into (II.2) finally gives

$$\begin{aligned}
&A_h \Pr\left(\sum_{i=2}^{n}\Lambda_i^2 \le r_{\Lambda_1}^2,\, \Lambda_2 \ge \beta_h(\Lambda_1),\, \Lambda_3 \ge -l_{w,h}(\Lambda_1,\Lambda_2)\right)\\
&\le \frac{A_h}{\sqrt{1+2t\eta}}\left(\frac{1}{\sqrt{1-2t}}\right)^{n-1}\exp\left\{\frac{\Delta_h^2\left(s-\frac{k\zeta_{w,h}}{\sqrt{1-\rho_{w,h}^2}}\right)^2}{2(1+2t\eta)}\right\}\\
&\quad\cdot\exp\left\{\frac{-4t\eta nc-2\sqrt{2nc}\left(s-\frac{k\zeta_{w,h}}{\sqrt{1-\rho_{w,h}^2}}\right)\Delta_h}{2(1+2t\eta)}\right\}\\
&\quad\cdot\exp\left\{\frac{\left(s-\frac{k\rho_{w,h}}{\sqrt{1-\rho_{w,h}^2}}\right)^2}{2(1-2t)}+\frac{k^2}{2(1-2t)}\right\}\\
&=\sqrt{\frac{1-2t}{1+2t\eta}}\,e^{-g(c,t,k,s,\eta,h,n)},\quad -\frac{1}{2\eta}<t\le 0,\ k\ge 0,\ s\ge 0
\end{aligned}$$

which proves the Chernoff bound in (I.9) with the function g introduced in (I.10).

APPENDIX III
MONOTONICITY W.R.T. THE CORRELATION COEFFICIENT

Consider the probabilities

$$\Pr(E_{0,i}(z_1),\, E_{0,j}^{\mathrm{c}}(z_1),\, \mathbf{y}\in C_n(\theta)\mid z_1)$$

and denote the Hamming weights of $\mathbf{c}_i$ and $\mathbf{c}_j$ by d_i and d_j, respectively. In [17], it is shown that as long as $d_i > d_j$, the probabilities above are monotonically decreasing functions of the correlation coefficients $\rho=\cos\phi$ between the noise components whose sizes are z_2 and z_3' (see the upper plot of Fig. 2). Hence, the optimization problem in (24) is simplified by choosing the first error event as well as the complementary error events on the RHS of (24) to correspond to a codeword with Hamming weight $d_{\min}$, and (25) is obtained. Here we prove that the aforementioned probabilities are monotonically decreasing functions of the correlation coefficients for *any* choice of i,j. As a consequence, one can obtain a version of the ITSB by setting in (24) $\pi_1=\lambda_i=w$ where $w\in\{d_{\min},\ldots,n\}$, and then choose the optimal w which minimizes the resulting upper bound. In order to prove this, we follow the steps in [17, Appendix I] where it is shown that the above probabilities are monotonically decreasing functions of ρ if

$$\frac{Z_2}{\beta_j(z_1)}>\rho. \tag{III.1}$$

Note that the joint event $(E_{0,i}(z_1),\ \mathbf{y}\in C_n(\theta))$ implies that the noise component Z_2 is in the range between $\beta_i(z_1)$ and r_{z_1} (see Fig. 1), so the minimal value of the RHS of (III.1) is

$$\frac{\beta_i(z_1)}{\beta_j(z_1)}=\sqrt{\frac{d_i(n-d_j)}{d_j(n-d_i)}}$$

where the last equality follows from (4) and since $\delta_h=2\sqrt{hE_{\mathrm{s}}}$ for BPSK modulated signals. Clearly,

$$\sqrt{\frac{d_i(n-d_j)}{d_j(n-d_i)}}>\frac{\min(d_i,d_j)[n-\max(d_i,d_j)]}{\sqrt{d_id_j(n-d_i)(n-d_j)}} \tag{III.2}$$

and from (33), it is evident that the RHS of (III.2) is the maximal value of ρ, so condition (III.1) is always satisfied while referring to the joint event $(E_{0,i}(z_1),\ \mathbf{y}\in C_n(\theta))$ given that $Z_1=z_1$.

REFERENCES

[1] 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.

[2] T. M. Duman and M. Salehi, "New performance bounds for turbo codes," *IEEE Trans. on Communications*, vol. 46, pp. 717–723, June 1998.

[3] T. M. Duman, *Turbo Codes and Turbo Coded Modulation Systems: Analysis and Performance Bounds*, Ph.D. dissertation, Elect. Comput. Eng. Dep., Northeastern University, Boston, MA, USA, May 1998.

[4] M. Fossorier, "Critical point for maximum-likelihood decoding of linear block codes," *IEEE Communications Letters*, vol. 9, no. 9, pp. 817–819, September 2005.

[5] J. Galambos and I. Simonelli, *Bonferroni-type inequalities with applications*, Springer Series in Statistics, Probability and its Applications, Springer-Verlag, New-York, 1996.

[6] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. on Information Theory*, vol. 11, pp. 3–18, January 1965.

[7] H. Herzberg and G. Poltyrev, "Techniques of bounding the probability of decoding error for block modulation structures," *IEEE trans. on Information Theory,* vol. 40, no. 3, pp. 903–911, May 1994.

[8] D. Hunter, "An upper bound for the probability of a union," *Journal of Applied Probability*, vol. 13, pp. 597–603, 1976.

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

[10] I. Sason and S. Shamai, "Bounds on the error probability of ML decoding for block and turbo-block codes," *Annals of Telecommunication*, vol. 54, no. 3–4, pp. 183–200, March–April 1999.

[11] I. Sason and S. Shamai, "Improved upper bounds on the ML decoding error probability of parallel and serial concatenated turbo codes via their ensemble distance spectrum," *IEEE Trans. on Information Theory*, vol. 46, no. 1, pp. 24–47, January 2000.

[12] I. Sason and S. Shamai, "Performance analysis of linear codes under maximum-likelihood decoding: a tutorial," *Foundations and Trends in Communications and Information Theory*, vol. 3, no. 1–2, pp. 1–222, NOW Publishers, Delft, the Netherlands, July 2006.

[13] N. Shulman and M. Feder, "Random coding techniques for nonrandom codes,"*IEEE Trans. on Information Theory*, vol. 45, no. 6, pp. 2101–2104, September 1999.

[14] M. Twitto, I. Sason and S. Shamai, "Tightened upper bounds on the ML decoding error probability of binary linear block codes," to appear in the *IEEE Trans. on Information Theory*, 2007.

[15] S. Yousefi and A. Khandani, "Generelized tangential sphere bound on the ML decoding error probability of linear binary block codes in AWGN interference," *IEEE Trans. on Information Theory*, vol. 50, no. 11, pp. 2810–2815, November 2004.

[16] S. Yousefi and A. K. Khandani, "A new upper bound on the ML decoding error probability of linear binary block codes in AWGN interference," *IEEE Trans. on Information Theory*, vol. 50, no. 12, pp. 3026–3036, December 2004.

[17] S. Yousefi and A. Mehrabian, "Improved tangential sphere bound on the ML decoding error probability of linear binary block codes in AWGN interference," *Proceedings* 37^{th} *Annual Conference on Information Science and Systems (CISS 2005)*, John Hopkins University, Baltimor, MD, USA, March 16–18, 2005.

[18] S. Yousefi, "Gallager first bounding technique for the performance evaluation of maximum-likelihood decoded linear binary block codes," *IEE Proceedings on Communications*, vol. 153, no. 3, pp. 317–332, June 2006.

[19] J. Zangl and R. Herzog, "Improved tangential sphere bound on the bit error probability of concatenated codes," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 5, pp. 825–837, May 2001.