Gallager Bounds for Linear Codes in Binary-Input Output-Symmetric Memoryless Channels

Alfonso Martinez
Technische Universiteit Eindhoven
Eindhoven, The Netherlands
e-mail: A.Martinez@tue.nl

Albert Guillén i Fàbregas
University of South Australia
Mawson Lakes, Australia
e-mail: albert.guillen@unisa.edu.au

Giuseppe Caire
Institut Eurécom
Sophia-Antipolis, France
e-mail: giuseppe.caire@eurecom.fr

***Abstract*— This paper presents a general methodology to extend Gallager bounds on the maximum-likelihood decoding error probability to arbitrary binary-input output-symmetric memoryless channels. Based on the log-likelihood ratios, a new space is constructed in which the signals naturally lie on a sphere, and for which geometric analysis is straightforward. In particular, we focus on Poltyrev's tangential-sphere bound, and we illustrate its connections with the Engdahl-Zigangirov bound. Approximations to these bounds are shown to be very tight.**

I. Introduction

The tightest upper bounds on the error probability under maximum likelihood (ML) decoding in the additive white Gaussian noise (AWGN) channel are based on a technique devised by Gallager [3] (we shall refer to these bounds as Gallager bounds in the following) and revisited by Shamai and Sason [4]. These bounds are often geometric in nature and exploit that the transmitted codewords lie on an n-dimensional sphere (for binary transmission) and that additive Gaussian noise is independently added along each axis.

In this paper, we present a new geometric construction for arbitrary binary-input output-symmetric (BiOS) memoryless channels, such as binary-input fading channels, or bit-interleaved coded modulation (BICM) systems which includes binary transmission over AWGN as a particular case. The construction creates an equivalent channel, the Λ-channel, which verifies that 1) the equivalent received codewords $\boldsymbol{\omega}^{(m)}$ lie on a sphere, even though they the real received codewords may not have the same energy (such as in fading channels), and 2) that the BiOS channel is transformed into an equivalent channel with additive non-Gaussian noise. We then show how this construction allows for an immediate extension of Gallager bounds based on geometric considerations and finally present a very tight approximation to the bounds.

II. Improved Bounds for the AWGN Channel

Consider first the following channel model,

$$z_i = w_i + \eta_i, \quad i = 1, \ldots, n \tag{1}$$

where $z_i \in \mathbf{R}$ is the received signal at the i-th instant, $w_i \in \{-1, +1\}$ is the BPSK transmitted symbol at instant i and $\eta_i \in \mathbf{R}$ is the corresponding noise sample assumed to be i.i.d. Gaussian. We denote the transmitted and received vectors $\mathbf{w} = (w_1, w_2, \ldots, w_n) \in \mathbf{R}^n$ and $\mathbf{z} = (z_1, z_2, \ldots, z_n) \in \mathbf{R}^n$. The vectors $\mathbf{w}$ have energy n and lie on a n-dimensional sphere of radius $\sqrt{n}$. We assume that $\mathbf{w}$ are the BPSK mapping of the codewords $\mathbf{c} = (c_1, c_2, \ldots, c_n)$ of a linear binary code of length n and dimension k.

Let codeword m be sent, and its corresponding BPSK vector be denoted by $\mathbf{w}^{(m)}$. The decoder selects the codeword $\hat{m}$ such that $\hat{m} = \arg\max_{m'} \Pr(\mathbf{w}^{(m')}|\mathbf{z})$. The error probability is then $\Pr_e(m) = \Pr(\hat{m} \neq m)$. Finally, for linear codes and equiprobable codewords, we have $\Pr_e = \Pr_e(m)$.

Most efficient bounds on $\Pr_e(m)$ are based on a basic technique devised by Fano and Gallager [3]. The space of possible received signals is partitioned in two disjoint subsets, named "good" and "bad", and denoted by $\mathcal{Z}_{\mathrm{G}}$ and $\mathcal{Z}_{\mathrm{B}}$ respectively. Then we have that[1],

$$\begin{aligned}\Pr_e(m) &= \Pr_e(\mathbf{z} \in \mathcal{Z}_{\mathrm{G}}) + \Pr_e(\mathbf{z} \in \mathcal{Z}_{\mathrm{B}}) \\ &\leq \Pr_e(\mathbf{z} \in \mathcal{Z}_{\mathrm{G}}) + \Pr(\mathbf{z} \in \mathcal{Z}_{\mathrm{B}}) \\ &\leq \Pr(\mathbf{z} \in \mathcal{Z}_{\mathrm{B}}) + \sum_{m' \neq m} \Pr_{\mathrm{e}}\big(m, m' | \mathbf{z} \in \mathcal{Z}_{\mathrm{G}}\big), \end{aligned} \tag{2}$$

where we have applied a union bound to sum over all $2^k - 1$ candidate codewords $m' \neq m$. We also denote the pairwise error probability $\Pr_{\mathrm{e}}(m, m')$ in the "good" region $\mathcal{Z}_{\mathrm{G}}$ by $\Pr_{\mathrm{e}}\big(m, m' | \mathbf{z} \in \mathcal{Z}_{\mathrm{G}}\big)$.

Recall that for AWGN, and assuming equally likely codewords, the pairwise error probability (dropping the conditioning $\mathbf{z} \in \mathcal{Z}_{\mathrm{G}}$ for simplicity) is obtained by comparing the Euclidean distance $\big|\mathbf{z} - \mathbf{w}\big|^2$ between the received signal and codewords m and m',

$$\Pr_{\mathrm{e}}(m, m') = \Pr\big(\Pr(\mathbf{w}^{(m')}|\mathbf{z}) > \Pr(\mathbf{w}^{(m)}|\mathbf{z})\big) \tag{3}$$

$$= \Pr\Big(\big|\mathbf{z} - \mathbf{w}^{(m')}\big|^2 < \big|\mathbf{z} - \mathbf{w}^{(m)}\big|^2\Big). \tag{4}$$

A. The Tangential Sphere Bound

1) Definition of "Good" and "Bad" Regions: We follow the original presentation by Poltyrev [1], and add where necessary some minor modifications. The region $\mathcal{Z}_{\mathrm{G}}$ is a cone[2] whose axis is the line connecting the origin of $\mathbf{R}^n$,

[1] Throughout the paper, we use $\Pr(\cdot)$ to refer to a probability in a general sense: it can be event probabilities, the distribution of discrete random variables, or the density of a real-valued random variable. This unifies notation, and the context makes it clear the precise meaning of $\Pr(\cdot)$.

[2] Differently from Poltyrev, we consider a single cone, rather than a double cone as he did; this seems more natural and general a choice.

$\mathbf{0} = (0, \ldots, 0)$ with the point $\mathbf{w}^{(m)}$. The cone has half-angle θ and the vertex of the cone is located at a distance ρ_0 from $\mathbf{w}^{(m)}$. Figure 1 depicts such a cone for the very simple case $n = 2$. The all-zero codeword $(-1, -1)$ is also depicted, as well as a candidate codeword at Hamming distance $h = 1$.

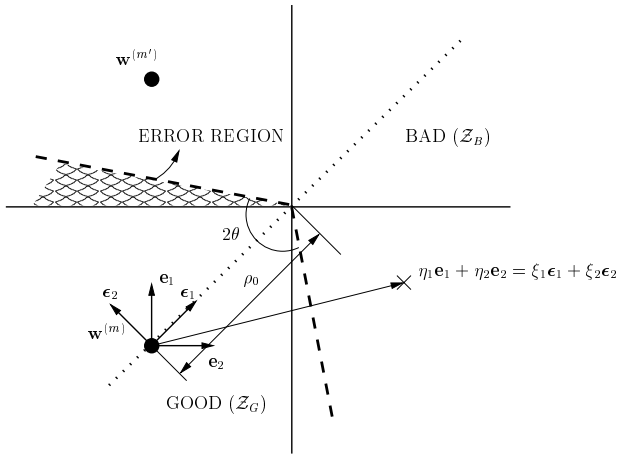

Fig. 1. Picture of a (single) cone in $\mathbf{R}^2$; it separates the space into two regions, $\mathcal{Z}_\mathrm{G}$ and $\mathcal{Z}_\mathrm{B}$. The error region in $\mathcal{Z}_\mathrm{G}$ is also depicted.

2) Change of Coordinates: Consider the *natural* basis of $\mathbf{R}^n$ be the vectors $\mathbf{e}_i = (\underbrace{0, \ldots, 0}_{i-1}, 1, \underbrace{0, \ldots, 0}_{n-i})$ for $i = 1, \ldots, n$. Instead of using this basis for the analysis, it is convenient to change to a new (orthonormal) basis with some useful properties. We denote such a *new* basis by $\boldsymbol{\epsilon}_i$, for $i = 1, \ldots, n$. Therefore, we have that $\boldsymbol{\epsilon}_i = \sum_{j=1}^n \alpha_{i,j}\mathbf{e}_j$, where the coefficients $\alpha_{i,j}$ define the change-of-basis matrix. The coefficients $\alpha_{i,j}$ are determined by the following conditions

1) The first new basis vector, $\boldsymbol{\epsilon}_1$, should lie along the axis of the cone, and should point to the origin. Then, we easily obtain that $\alpha_{1,j} = \frac{1}{\sqrt{n}}, 1 \le j \le n$. Note that it is independent of m'.
2) The second new basis vector lies on the plane determined by $\mathbf{0}$, $\mathbf{w}^{(m)}$ and $\mathbf{w}^{(m')}$, and is orthogonal to $\boldsymbol{\epsilon}_1$. An application of the Gram-Schmidt orthogonalization procedure gives

$$\alpha_{2,j} = \begin{cases} \frac{1}{\sqrt{n}}\sqrt{\frac{n-h}{h}}, & 1 \le j \le h, \\ -\frac{1}{\sqrt{n}}\sqrt{\frac{h}{n-h}}, & h+1 \le j \le n. \end{cases}$$

where h is the Hamming distance between $\mathbf{w}^{(m)}$ and $\mathbf{w}^{(m')}$. The basis vectors $\mathbf{e}_1, \mathbf{e}_2$ and $\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2$ are depicted in Figure 1 once translated onto the transmitted codeword $\mathbf{w}^{(m)}$. The remaining $n-2$ coordinates, $\boldsymbol{\epsilon}_i$ for $i > 2$, orthogonal to the first two, are irrelevant for the calculation of the ML error between codewords m and m'. The noise vector has now different coordinates in the new basis, denoted by ξ_i, and given by $\xi_i = \sum_{j=1}^n \alpha_{i,j}\eta_j$; the distribution of ξ_i remains i.i.d. Gaussian [5].

The points of the cone with fixed ξ_1 lie on an $(n-1)$-dimensional sphere and satisfy $\sum_{i=2}^n \xi_i^2 = \rho^2(\xi_1)$, where $\rho(\xi_1)$ is the radius of the sphere. The radius is given by

$$\rho(\xi_1) = \begin{cases} 0, & \xi_1 \ge \rho_0 \\ (\rho_0 - \xi_1)\tan\theta, & -\infty < \xi_1 \le \rho_0. \end{cases}$$

3) Calculation of Probabilities: In terms of the new coordinates, an error is made when $\sqrt{n}\sin\phi \le \xi_1\sin\phi + \xi_2\cos\phi$, where ϕ is half the angle between the vectors $\mathbf{w}^{(m)}$ and $\mathbf{w}^{(m')}$,

$$\cos 2\phi = \frac{\langle \mathbf{w}^{(m)}, \mathbf{w}^{(m')} \rangle}{|\mathbf{w}^{(m)}|\,|\mathbf{w}^{(m)}|} = \frac{n-2h}{n} \Longrightarrow \tan\phi = \left(\frac{h}{n-h}\right)^{1/2}.$$

Defining $\beta_h(\xi_1) = (\sqrt{n} - \xi_1)\tan\phi$, the condition for error within the cone corresponds to $\beta_h(\xi_1) \le \xi_2 \le \rho(\xi_1)$.

We now invoke the union bound Eq. (2) and calculate the contribution from the "good" region as[3]

$$\Pr_e(\mathbf{z} \in \mathcal{Z}_\mathrm{G}) \le \sum_h A_h(\theta) \int_{-\infty}^{\rho_0} \int_{\beta_h(\xi_1)}^{\rho(\xi_1)} \Pr(\xi_1, \xi_2) d\xi_2\, d\xi_1 \quad (5)$$

$$\Pr(\xi_1, \xi_2) = \Pr\left(\sum_{i=3}^n \xi_i^2 \le \left(\rho^2(\xi_1) - \xi_2^2\right)\right) \Pr(\xi_1)\Pr(\xi_2).$$

We have exploited the linearity of the code to define the number of codewords with Hamming weight h by A_h, $1 \le h \le n$ and group the terms with common Hamming distance in the bound. Of these, we need consider only the terms with h such that $\tan\phi = \sqrt{\frac{h}{n-h}} < \tan\theta$, that is, $A_h(\theta) = A_h$ if $\tan\phi < \tan\theta$, and zero otherwise. The independence of the transformed variables ξ_i allows us to factorize the joint densities into the product of marginal densities.

Similarly we obtain for the "bad" region

$$\Pr(\mathbf{z} \in \mathcal{Z}_\mathrm{B}) = \int_{-\infty}^{\rho_0} \Pr\left(\sum_{i=2}^n \xi_i^2 > \rho^2(\xi_1)\right) \Pr(\xi_1)\, d\xi_1 + \int_{\rho_0}^{+\infty} \Pr(\xi_1)\, d\xi_1. \quad (6)$$

4) Solution of the Optimization Problem: The standard TSB sets $\rho_0 = \sqrt{n}$. Then, the optimum value of θ, is found by setting the partial derivatives to zero. With some algebra, the optimality condition can be expressed as

$$\sum_h A_h(\theta) \frac{\Gamma\left(\frac{1}{2}(n-1)\right)}{\sqrt{\pi}\,\Gamma\left(\frac{1}{2}(n-2)\right)} \int_0^{\varphi_h} \sin\varphi^{n-3}\, d\varphi = 1, \quad \text{where}$$

$$\varphi_h = \arccos\frac{\tan\phi}{\tan\theta} = \arccos\left(\frac{1}{\tan\theta}\sqrt{\frac{h}{n-h}}\right). \quad (7)$$

B. Connection with the Engdahl-Zigangirov Bound

1) "Good" and "Bad" Regions: The general optimization program that minimizes Eqs. (5)-(6) includes two variables ρ_0 and θ. The standard formulation of the TSB sets ρ_0 and optimizes θ. An interesting possibility is to fix θ and optimize the cone apex ρ_0. Setting the angle $\theta = \pi/2$ makes the cone

[3]Although this analysis holds for block error probability, it is very easy to show that the same bounding technique can be applied for bit error probabilities, if A_h is given a suitable meaning. Similarly, the bound is also applicable to average performance for ensembles of codes.

collapse into a half-space, and the axis of the cone to a single point. This is the "good" region considered by Engdahl and Zigangirov [2] for their analysis of binary AWGN. If we set $\tan\theta = +\infty$, the radius of the cone is

$$\rho(\xi_1) = \begin{cases} 0, & \xi_1 < \rho_0 \\ +\infty, & -\infty < \xi_1 \le \rho_0. \end{cases}$$

which implies that $A_h(\theta) = A_h$ for every h. We can now rewrite Eqs. (5)-(6) as follows:

$$\begin{aligned} \Pr_e(\mathbf{z} \in \mathcal{Z}_{\mathrm{G}}) &= \sum_h A_h \int_{-\infty}^{\rho_0} \int_{\beta_h(\xi_1)}^{+\infty} \Pr(\xi_1)\Pr(\xi_2)\, d\xi_2\, d\xi_1 \\ &= \sum_h A_h \int_{-\infty}^{\rho_0} Q\Big(\frac{\beta_h(\xi_1)}{\sigma}\Big) \Pr(\xi_1)\, d\xi_1, \qquad (8) \\ \Pr(\mathbf{z} \in \mathcal{Z}_{\mathrm{B}}) &= \int_{\rho_0}^{+\infty} \Pr(\xi_1)\, d\xi_1. \qquad (9) \end{aligned}$$

2) Solution of the Optimization Problem: The optimization of Eqs. (8), (9) is done by setting the first derivative with respect to ρ_0 to zero. We obtain

$$\begin{aligned} &\sum_h A_h Q\Big(\frac{\beta_h(\rho_0)}{\sigma}\Big) \Pr(\rho_0) - \Pr(\rho_0) = 0 \\ &\qquad \Longrightarrow \sum_h A_h Q\Big(\frac{\beta_h(\rho_0)}{\sigma}\Big) = 1. \end{aligned}$$

This is equivalent to the same optimization equation in [2], after a change of variable which does not modify the bound.

III. Extension to BiOS Memoryless Channels

A. A Posteriori Log-Likelihood Ratio

For the purpose of estimating the error probability, the direct use the received signal $\mathbf{z}$ in Eq. (3) can be circumvented. We shall show that it is in fact sufficient to consider a different variable, the *a posteriori log-likelihood ratio*, denoted by $\Lambda_i(c)$, and given by

$$\Lambda_i(c) = \log \frac{\Pr(\hat{c}_i = \bar{c} | \mathcal{V}_i(c))}{\Pr(\hat{c}_i = c | \mathcal{V}_i(c))} \qquad (10)$$

where $\hat{c}_i$ is the i-th bit, c is the sent bit (assumed known for decoding error estimation), $\bar{c}$ its complement. $\mathcal{V}_i(c)$ is a vector containing all the random elements in the channel when bit c is sent[4] at time i. Note that this definition is slightly different from that of the standard log-likelihood ratio, as the bit at denominator is the trasmitted bit c rather than 0. We shall see later that this convention brings about new clarity in the geometric analysis.

It is an easy consequence of the definition of BiOS channels that the distribution of $\Lambda(c)$ is actually independent of the transmitted bit (be it 0 or 1), as shown by

[4]In the case of AWGN it is only the noise realization. In the binary-input Rayleigh fading channel, $\mathcal{V}$ includes both, fading (in the form of channel state information) and noise realizations. In the case of BICM, $\mathcal{V}_i$ includes the noise, fading (if any), the signal constellation and the binary label [6], [7]. It may also include the effects of incoherent detection, or imperfect channel estimation.

Proposition 1 *For memoryless BiOS channels $\Lambda_i(c)$ are i.i.d. and, in particular, are identical for both values of the transmitted bit, i. e., $c = 0$ and $c = 1$.*

Proof: Thanks to the absence of memory in the channel we may safely drop the time index i. Let us assume that bit c was transmitted. The channel output z is distributed as $\Pr(z|c)$. The density of $\Lambda(c)$ is given by the theorem of total probability [5] by integrating over all possible channel outputs z

$$\Pr\big(\Lambda(c) = \lambda\big) = \int_{z|\Lambda(c)=\lambda} \Pr(z|c)\, \mu(dz).$$

Here $\mu(dz)$ is the measure of the interval dz. Inverting the sign of the outputs and using the definition of output-symmetric channels ($\Pr(z|c) = \Pr(-z|\bar{c})$) we have

$$\begin{aligned} \Pr\big(\Lambda(c) = \lambda\big) &= \int_{z|\Lambda(c)=\lambda} \Pr(-z|c)\, \mu(dz) \\ &= \int_{z|\Lambda(\bar{c})=\lambda} \Pr(z|\bar{c})\, \mu(dz) = \Pr\big(\Lambda(\bar{c}) = \lambda\big). \end{aligned}$$

■

This result allows us to drop the bit value c in the definition of Λ and in its density function, which we denote by $\Pr(\Lambda)$.

In the AWGN channel, it is straightforward to show that $\Lambda \sim \mathcal{N}(-4\,\mathrm{SNR}, 8\,\mathrm{SNR})$. In the case of BPSK transmission over the fully-interleaved Rayleigh fading channel, we have that conditioned on the instantaneous fading $\gamma \in \mathbf{C}$, $\Lambda \sim \mathcal{N}(-4|\gamma|^2\,\mathrm{SNR}, 8|\gamma|^2\,\mathrm{SNR})$. If we let $\chi = |\gamma|^2$, then χ is then exponentially distributed with density $\Pr_\chi(\chi) = \frac{1}{2\sigma^2} e^{-\frac{\chi}{2\sigma^2}}$ where $\sigma^2 = 1$. This integral takes the closed form

$$\Pr(\Lambda) = \frac{1}{4\sqrt{\mathrm{SNR} + \mathrm{SNR}^2}} \exp\Big(-\frac{\Lambda}{2} - \frac{|\Lambda|}{2}\sqrt{\frac{1+\mathrm{SNR}}{\mathrm{SNR}}}\Big),$$

i. e., a two-sided exponential with mean $\langle\Lambda\rangle = -4\,\mathrm{SNR}$ and variance $\mathrm{Var}(\Lambda) = 8\,\mathrm{SNR} + 16\,\mathrm{SNR}^2 = 8\,\mathrm{SNR}(1 + 2\,\mathrm{SNR})$. For BICM [6], the codewords of $\mathcal{C}$ are bit-interleaved and mapped onto constellation symbols $w \in \mathbf{C}$. After symmetrizing the bit labeling (if required) [6], BICM is another example of BiOS channel. In this case Λ does not have a simple distribution but its moments can be easily evaluated [7].

B. Λ-Sphere and Λ-Channel

In this section, we introduce another set of variables based on Λ which gives the same result for the AWGN channel and enables a formulation of Gallager bounds for general BiOS memoryless channels.

The first step is to define the channel codeword . To every codeword, say m, we associate a point $\boldsymbol{\omega}^{(m)} \in \mathbf{R}^n$,

$$\boldsymbol{\omega}^{(m)} \triangleq r\mathbf{w}^{(m)}, \qquad (11)$$

where $\mathbf{w}^{(m)}$ is the BPSK mapping of the codeword, $r \triangleq \big|\langle\Lambda\rangle\big|$ and $\langle\Lambda\rangle \triangleq \mathrm{E}[\Lambda] \le 0$. Eq. (11) constitutes the core of our geometric approach: the points $\boldsymbol{\omega}^{(m)}$ all lie now on the n-dimensional sphere in $\mathbf{R}^n$ with center at the origin and radius $r\sqrt{n}$. Along each axis of $\mathbf{R}^n$ we define an equivalent additive

zero-mean noise $\tilde{\eta}_i$ of value $\tilde{\eta}_i \stackrel{\Delta}{=} \Lambda_i - \langle\Lambda\rangle$ which results in the following Λ-channel,

$$\Lambda_i = -r + \tilde{\eta}_i.$$

For AWGN, the noise distribution is $\tilde{\eta}_i \sim \mathcal{N}(0, 8\,\mathrm{SNR})$ and $r = 4\,\mathrm{SNR}$. After dividing all variables Λ_i by a common factor r, the normalized variables points $\boldsymbol{\omega}^{(m)}/r$ fall back on a sphere of radius $\sqrt{n}$, and the movements along each coordinate axis are Gaussian random variables with zero mean and variance $\sigma^2 = 8\,\mathrm{SNR}/r^2 = 1/(2\,\mathrm{SNR})$. We recover thus the original geometric problem, as in Eq.(1), as expected.

C. Pairwise Error Probability in the Λ-Channel

Consider now the transmission of two codewords, m and m', at Hamming distance h. Let $\boldsymbol{\Lambda} = (\Lambda_1, \ldots, \Lambda_n)$ denote the vector of all realizations of Λ_i. The following theorem is the key that allows us to construct geometric Gallager bounds in the Λ-channel[5]

Theorem 1 *The pairwise error probability is a function of the Euclidean distance in the space of* $\boldsymbol{\Lambda}$*,*

$$\mathrm{Pr}_e(m, m') = \mathrm{Pr}\Big(\big|\boldsymbol{\Lambda} - \boldsymbol{\omega}^{(m')}\big|^2 < \big|\boldsymbol{\Lambda} - \boldsymbol{\omega}^{(m)}\big|^2\Big).$$

Proof: We start the analysis at Eq. (2), which links the a posteriori probabilities. The dependence on $\mathbf{z} \in \mathcal{Z}_G$ is irrelevant, so we safely drop it. Taking logarithms, and exploiting the absence of memory in the channel, we rewrite the condition for error (3) as

$$\begin{aligned}\mathrm{Pr_e}(m, m') &= \mathrm{Pr}\left(\sum_{i=1}^{n} \log \frac{\mathrm{Pr}\big(w_i^{(m')}|z_i\big)}{\mathrm{Pr}\big(w_i^{(m)}|z_i\big)} > 0\right) \\ &= \mathrm{Pr}\left(\sum_{i=1}^{n} \tilde{\Lambda}_i > 0\right),\end{aligned}$$

where $\tilde{\Lambda}_i \stackrel{\Delta}{=} \log \frac{\mathrm{Pr}\left(w_i^{(m')}|z_i\right)}{\mathrm{Pr}\left(w_i^{(m)}|z_i\right)}$. Note that its value is different from zero only for the h positions where $w_i^{(m)}$ and $w_i^{(m')}$ differ. By Proposition 1, the distribution of Λ is independent of the actual value of c. This allows us to ignore the value of $w_i^{(m)}$, and write

$$\mathrm{Pr_e}(m, m') = \mathrm{Pr}\left(\sum_{i=1}^{h} \Lambda_i > 0\right). \tag{12}$$

The difference in Euclidean distances, Δ, is

$$\Delta = \big|\boldsymbol{\Lambda} - \boldsymbol{\omega}^{(m')}\big|^2 - \big|\boldsymbol{\Lambda} - \boldsymbol{\omega}^{(m)}\big|^2. \tag{13}$$

As $\big|\omega_i^{(m)}\big|$ is constant, common terms cancel out and we have

$$\begin{aligned}\Delta &= -2\left(\sum_{i=1}^{n} \Lambda_i \left(\omega_i^{(m')} - \omega_i^{(m)}\right)\right) \\ &= -2r\left(\sum_{i=1}^{n} \Lambda_i \left(w_i^{(m')} - w_i^{(m)}\right)\right) = -4r\left(\sum_{i=1}^{h} \Lambda_i\right).\end{aligned}$$

From Eq. (12) this is equivalent to $\mathrm{Pr_e}(m, m') = \mathrm{Pr}(\Delta < 0)$. ■

D. Gallager Bounds for Linear Codes in the Λ-Channel

The previous theorem implies that we can partition the set of possible $\boldsymbol{\Lambda}$, $\mathbf{R}^n$, into two disjoint, "good" and "bad" regions, respectively denoted by $\mathcal{L}_\mathrm{G}$ and $\mathcal{L}_\mathrm{B}$, for arbitrary memoryless BiOS channels. We obtain then

Theorem 2 *The error probability for codeword* m *is upper bounded* [6] *by*

$$\mathrm{Pr}_e(m) \le \mathrm{Pr}(\boldsymbol{\Lambda} \in \mathcal{L}_B) + \sum_{m' \neq m} \mathrm{Pr}\left(\sum_{i=1}^{h} \Lambda_i > 0, \boldsymbol{\Lambda} \in \mathcal{L}_G\right).$$

Following the basic steps of the derivation presented in section II-A, instead of working along the natural basis of $\mathbf{R}^n$, we change basis to align it with the cone axis. With slight abuse of notation, let ξ_i be the noise vector in the transformed coordinates, $\xi_i = \sum_{j=1}^{n} \alpha_{i,j}\tilde{\eta}_j$. Therefore, the first two elements are

$$\begin{aligned}\xi_1 &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\big(\Lambda_j - \langle\Lambda\rangle\big) = -\sqrt{n}\,\langle\Lambda\rangle + \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\Lambda_j \\ \xi_2 &= \frac{1}{\sqrt{n}}\left(\sqrt{\tfrac{n-h}{h}}\sum_{j=1}^{h}\Lambda_j - \sqrt{\tfrac{h}{n-h}}\sum_{j=h+1}^{n}\Lambda_j\right).\end{aligned}$$

For later use, let us define a variable $\Sigma = \sum_{j=2}^{n}\xi^2$. The invariance of $\sum_{j=1}^{n}\tilde{\eta}_j^2$ to changes of coordinates makes it rather easy to see that its value is also given by $\Sigma = \sum_{j=1}^{n}\tilde{\eta}_j^2 - \xi_1^2$.

Applying Theorem 2 and the linearity of the code we can re-write (5)-(6) to obtain the desired expression for the TSB

$$\begin{aligned}\mathrm{Pr}_e(\boldsymbol{\Lambda} \in \mathcal{L}_\mathrm{G}) &= \sum_h A_h(\theta)\int_{-\infty}^{\rho_0}\int_{\beta_h(\xi_1)}^{\rho(\xi_1)} \mathrm{Pr}(\xi_1, \xi_2)\,d\xi_2\,d\xi_1 \\ \mathrm{Pr}(\xi_1, \xi_2) &= \mathrm{Pr}\Big(\sum_{i=3}^{n}\xi_i^2 \le \big(\rho^2(\xi_1) - \xi_2^2\big), \xi_1, \xi_2\Big) \\ \mathrm{Pr}(\boldsymbol{\Lambda} \in \mathcal{L}_\mathrm{B}) &= \int_{-\infty}^{\rho_0}\mathrm{Pr}\Big(\sum_{i=2}^{n}\xi_i^2 > \rho^2(\xi_1), \xi_1\Big)\,d\xi_1 \\ &\quad + \int_{\rho_0}^{+\infty}\mathrm{Pr}(\xi_1)\,d\xi_1.\end{aligned}$$

Recalling Eqs. (8) and (9), the EZB takes the form [7]

$$\begin{aligned}\mathrm{Pr}_e(\boldsymbol{\Lambda} \in \mathcal{L}_\mathrm{G}) &= \sum_h A_h \int_{-\infty}^{\rho_0}\mathrm{Pr}\big(\xi_2 > \beta_h(\xi_1), \xi_1\big)\,d\xi_1, \\ \mathrm{Pr}(\boldsymbol{\Lambda} \in \mathcal{L}_\mathrm{B}) &= \int_{\rho_0}^{+\infty}\mathrm{Pr}(\xi_1)\,d\xi_1.\end{aligned}$$

Note that, differently from the Gaussian case, the joint densities cannot be separated here. Fortunately, even though

[5] In general Λ is real-valued, but it may be discrete. In this case, it is important to track the probability $\mathrm{Pr}(\Lambda = 0)$, which may be non-zero, and account for in Eq. (12).

[6] Once again we keep implicit the term linked to $\mathrm{Pr}(\Lambda = 0)$.

[7] It is worthwhile noting that Sason [9] reached an identical generalization for the EZB in the binary fully-interleaved fading channel. However, his reasoning was based on correlations between sequences $\mathbf{z}$ and $\mathbf{w}$, thus of more difficult generalization for arbitrary channels.

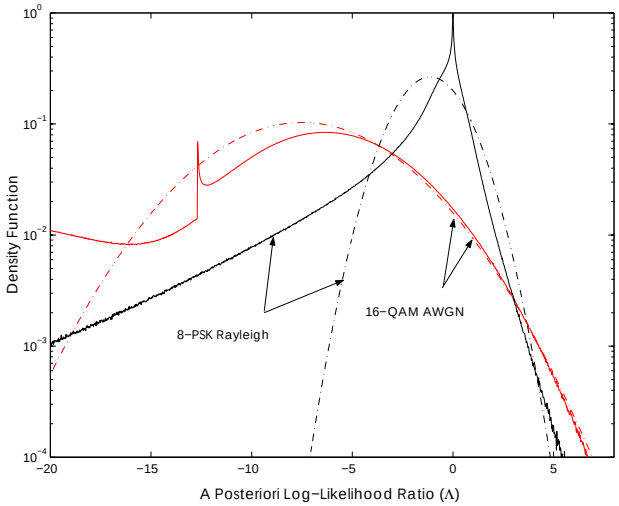

Fig. 2. Density $\Pr(\Lambda)$ for 16-QAM/8-PSK with Gray mapping over AWGN/Rayleigh fading and corresponding Gaussian approximations.

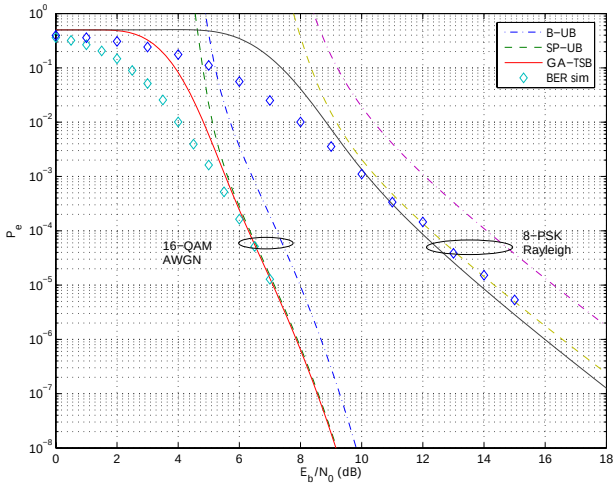

Fig. 3. BER simulations and bounds for BICM with 16-QAM/8-PSK with Gray mapping over AWGN/Rayleigh fading channels with optimal rate $\frac{1}{2}$ / $\frac{2}{3}$ convolutional codes respectively.

the variables are not independent, they are uncorrelated, and furthermore their dependence is extremely weak, due to the large number of variables involved in the transformations.

For the general case, there are no closed-form expressions for the tail probabilities we wish to calculate. In the AWGN case, they were normal, or chi-square distributed, with known formulas for the tail probabilities. In our case we must resort to saddlepoint approximations [8], which are very accurate and not difficult to compute.

As example of the accuracy achievable by this geometric approach combined with saddlepoint approximations, in [7], [10], a Gaussian approximation to TSB on the error probability of BICM (in AWGN and fully interleaved Rayleigh fading) was presented and computed. Independence between ξ_1, ξ_2, and Σ was assumed, as well as a Gaussian density to the first two and a χ^2 density of the latter. Fig. 2 shows the density $\Pr(\Lambda)$ for BICM in AWGN and Rayleigh fading, with the corresponding Gaussian approximations. A simple application to the central limit theorem [5] to ξ_1 shows that its distribution tends towards a Gaussian variable for large values of n. This observation holds only partially for ξ_2, for which the Hamming distance among codewords is important. Finally, even though a χ^2 approximation to Σ is reasonable, it must be applied with care, especially if the higher order moments of the Λ_i are nonzero.

Fig. 3 shows the bit-error probability Bhattacharyya bound (B-UB), the saddlepoint approximation union bound (SP-UB) [7], [10], the TSB with the Gaussian approximation (GA-TSB) and simulation for BICM in AWGN and Rayleigh fading with convolutional codes. While in AWGN the Gaussian approximation is very accurate, in the case of fading it is slightly optimistic.

IV. Conclusions

A careful examination of the proof of Poltyrev's tangential sphere bound shows that it can be extended to arbitrary binary-input output-symmetric channels. It is shown that, for the purpose of error probability estimation and without information loss, the BiOS channel can be transformed into an equivalent channel with additive non-Gaussian noise. In a well-defined sense, the a posteriori log-likelihood ratios of the received signals lie on a sphere, and furthermore, the additivity property of AWGN remains valid. This implies that geometric analysis is possible, and the equations for generalized tangential sphere bound and Engdahl-Zigangirov bounds are provided. Finally, tight approximations to these bounds are presented and computed.

References

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

[2] K. Engdahl and K. S. Zigangirov, "Tighter bounds on the error probability of fixed convolutional codes," *IEEE Trans. Inform. Theory*, vol. 47, no. 4, pp. 1625–1630, May 2001.

[3] R. G. Gallager, *Low-Density Parity-Check Codes*. Cambridge (USA): MIT Press, 1963.

[4] S. Shamai (Shitz) and I. Sason, "Variations on the Gallager bounds, connections and applications," *IEEE Trans. Inform. Theory*, vol. 48, no. 12, pp. 3029–3051, Dec. 2002.

[5] W. Feller, *An Introduction to Probability Theory and Its Applications*, 2nd ed. John Wiley & Sons, 1971, vol. 2.

[6] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inform. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.

[7] A. Martinez, A. Guillén i Fàbregas, and G. Caire, "New simple evaluation of the error probability of bit-interleaved coded modulation using the saddlepoint approximation," in *Proc. 2004 Int. Symp. on Information Theory and its Applications (ISITA 2004)*, Parma (Italy), Oct. 2004.

[8] J. L. Jensen, *Saddlepoint Approximations*. Oxford, UK: Clarendon Press, 1995.

[9] I. Sason and S. Shamai (Shitz), "On improved upper bounds on the decoding error probability of block codes over interleaved fading channels, with applications to turbo-like codes," *IEEE Trans. Inform. Theory*, vol. 47, no. 6, pp. 2275–2299, Sep. 2001.

[10] A. Martinez, A. Guillén i Fàbregas, and G. Caire, "Error probability analysis of bit-interleaved coded modulation," *submitted to IEEE Trans. Inform. Theory*, Nov. 2004.