IRWIN AND JOAN JACOBS
CENTER FOR COMMUNICATION AND INFORMATION TECHNOLOGIES

Gallager-Type Bounds for Non-Binary Linear Block Codes over Memoryless Symmetric Channels

Eran Hof, Igal Sason, Shlomo Shamai

CCIT Report #696
April 2008

DEPARTMENT OF ELECTRICAL ENGINEERING
TECHNION - ISRAEL INSTITUTE OF TECHNOLOGY, HAIFA 32000, ISRAEL

Gallager-Type Bounds for Non-Binary Linear Block Codes over Memoryless Symmetric Channels

CCIT Report #696 April 2008

Eran Hof Igal Sason Shlomo Shamai

Department of Electrical Engineering
Technion – Israel Institute of Technology
Haifa 32000, Israel
E-mails: {hof@tx, sason@ee, sshlomo@ee}.technion.ac.il

Abstract

The performance of non-binary linear block codes is studied in this paper via the derivation of new upper bounds on the error probability under ML decoding. The transmission of these codes is assumed to take place over a memoryless and symmetric channel. The new bounds, which rely on the Gallager bounding technique, are applied to expurgated ensembles of non-binary and regular low-density parity-check (LDPC) codes. These upper bounds are also compared with classical and recent sphere-packing lower bounds. This study indicates that the new upper bounds are useful for the performance evaluation of coded communication systems which incorporate non-binary coding techniques.

Index Terms

Block codes, linear codes, low-density parity-check (LDPC) codes, ML decoding, non-binary codes, sphere-packing bounds.

I. Introduction

The performance of coded communication systems is usually analyzed via upper and lower bounds on the decoding error probability. These bounds are of interest since the performance analysis of coded communication systems rarely admits exact expressions. Modern coding schemes (e.g., codes defined on graphs) perform reliably at rates which are close to the channel capacity, whereas union bounds are useless for codes of moderate to large block lengths at rates above the channel cut-off rate. The limitation of the union bound therefore motivates the introduction of some improved bounding techniques which can be also efficiently calculated. Although the performance analysis of specific codes is in general prohibitively complex, this kind of analysis is tractable for various code ensembles for which the derivation of some of their basic features (e.g., the average distance spectrum) lends itself to analysis. For a comprehensive tutorial on the performance analysis of binary linear block codes under maximum-likelihood (ML) decoding, the reader is referred to [1] and references therein, whereas this work is focused on non-binary linear block codes.

The 1965 Gallager bound [2] is one of the well-known upper bounds on the decoding error probability of ensembles of fully random block codes, and it is informative at all rates below the channel capacity limit. Emerging from the 1965 Gallager bound, the bounds of Duman and Salehi (see [3] and [4]) possess the pleasing feature that they are amenable to analysis for specific codes and general code ensembles for which the (average) distance spectrum analysis is tractable.

The framework of the Duman and Salehi bounding technique, in particular its second version (called hereafter the 'DS2 bound'), is generalized in [1], [5] and [6] for various memoryless communication systems. Moreover, this bound facilitates the derivation of a large class of previously reported bounds (or their Chernoff versions), as

This work will be presented in part at the 2008 IEEE Information Theory Workshop (ITW 2008) which will take place on May 5–9, 2008, in Porto, Portugal. This research work was supported by the Israel Science Foundation (grant no. 1070/07), and by the European Commission in the framework of the FP7 Network of Excellence in Wireless Communications NEWCOM++.

Igal Sason is the corresponding author (E-mail: sason@ee.technion.ac.il).

shown in [1] and [5]. Gallager-based bounds for binary linear block codes whose communication takes place over fading channels are provided in [7], [8] and [9]. The Shulman and Feder bound (SFB) [10] is another Gallager-type bound which coincides with the 1965 Gallager bound for fully random block codes and it can also be applied to structured codes or ensembles. An adaptation of the SFB to non-binary linear block codes was reported in [11] for the case of coding with a random coset mechanism (see, e.g., [11]–[14]), and for the case of transmission over modulo-additive noise channels (see [15]). Generalization of Gallager-type bounds, among them the DS2 bound, for the case of binary linear block codes whose transmission take places over parallel channels are provided in [16] and [17].

The 1959 sphere-packing (SP59) bound of Shannon [18] is a lower bound on the decoding error probability of block codes whose transmission takes place over the additive white Gaussian noise (AWGN) channel with equal-energy signaling. The 1967 sphere-packing bound of Shannon, Gallager and Berlakamp [19], forms another lower bound on the decoding error probability of block codes which applies to discrete memoryless channels. This bound was revisited by Valembois and Fossorier in [20] showing a remarkable improvement for finite-length block codes over the bound in [19] while extending its application to memoryless binary-input output-symmetric (MBIOS) channels. A comprehensive overview of classical sphere-packing bounds is provided in [1, Chapter 5]. An improved sphere-packing (ISP) bound, which holds for all memoryless symmetric channels, was recently derived in [21] by improving the bounding techniques in [19] and [20].

Low-density parity-check (LDPC) codes were proposed by Gallager in his seminal work [22]. The performance analysis of the binary LDPC ensembles in [22] is carried under the assumption that the communication channel is an MBIOS channel. In sharp contrast to the binary case, the performance analysis of non-binary LDPC code ensembles in [22] is carried under a symmetry assumption which is tailored to the specific bounding technique introduced in [22]. The asymptotic error performance of several non-binary LDPC structures is studied in [11] under ML decoding (the iterative performance of these structures is studied in [12], further bounds on the decoding threshold of non-binary LDPC ensembles under iterative decoding are studied in [23] and [24]). It is assumed in [11] that the transmission takes place over channels with a random coding mechanism. The random coding mechanism enables to dismiss the exhaustive symmetry condition required in [22]. The decoding error probability of various non-binary LDPC code constructions is studied empirically in [25] and [26]. Except for non-binary LDPC codes, turbo codes were also considered for high spectral efficiency schemes (see e.g., [27]-[31] and references therein).

The drawback of the union bound in over-counting points in the decision regions is pronounced for efficient moderate to long block code ensembles. This motivates this study in this paper which is focused on the derivation of upper bounds on the ML decoding error probability of ensembles of non-binary linear block codes whose transmission takes place over a memoryless symmetric channel. Our definition of symmetry for channels whose input is non-binary generalizes the common definition for MBIOS channels. Under these symmetry requirements, we prove that the conditional error probability under ML decoding is independent of the transmitted codeword. The general bounding approach used in this paper is based on a partitioning of the original ensemble into two subsets of codebooks according to their minimal Hamming weight. For the set of codebooks whose minimal distances are below a certain value (which is later determined in order to achieve a tight bound), a simple union bound is used which only depends on the distance properties of this considered set of codebooks. As for the other set of codebooks (whose minimal Hamming weight is larger than the above value), a Gallager-type bound on the decoding error probability is used; the latter bound depends both on the distance properties of the ensemble and the communication channel, and it relies on a generalization of the DS2 bound to non-binary linear block code ensembles.

The upper bounds on the error performance derived in this paper are applied to non-binary regular LDPC ensembles of Gallager [22]. The error performance is studied for various communication channel models: The q-ary symmetric channel, the AWGN channel with a q-ary phase shift keying (PSK) modulation, and fully interleaved fading channels with perfect channel state information (CSI) at the receiver. For the considered fading channels, a sub-optimal variation of the DS2 bound is derived which generalizes the results in [9] in terms of the alphabet cardinality. In addition, the derived upper bounds are compared to the SP59 and ISP lower bounds on the decoding error probability. The exact complete composition spectra for non-binary regular LDPC code ensembles is also provided (instead of the upper bound which is provided in [22]). The evaluation of the exact complete composition spectra of these non-binary and regular LDPC code ensembles forms a generalization of the analysis in [32].

This paper is structured as follows: The symmetry requirements and the message independence proposition are provided in Section II. The proposed bounding approach is introduced in Section III, and these bounds are

exemplified in Section III-C for the Gallager ensembles of non-binary regular LDPC codes whose transmission takes place over a q-ary symmetric channel and an AWGN channel with q-ary PSK modulation using the exact complete composition spectrum. Variations of the proposed bounds for fully-interleaved fading channels with perfect CSI at the receiver are studied and exemplified for the considered ensembles in Section IV. Section V concludes our discussion. Various technical details are relegated to the appendices.

II. Symmetry and Message Independence

Let $\mathcal{X} = \{x_0, x_1, \ldots, x_{q-1}\}$ be a given alphabet with a cardinality q. We assume an addition operation $(+)$ over the alphabet $\mathcal{X}$ for which $\{\mathcal{X}, +\}$ forms an Abelian group. Let $x_0 = 0$ be the additive identity of this group. In addition, let $\mathcal{Y}$ be a given discrete (or continuous) alphabet. We assume a memoryless channel, and denote the channel transition probability (or probability density, respectively) function by $p(y|x)$, where $x \in \mathcal{X}$ and $y \in \mathcal{Y}$.

Definition 1. A memoryless channel which is characterized by a transition probability p, an input-alphabet $\mathcal{X}$ and a discrete output alphabet $\mathcal{Y}$ is *symmetric* if there exists a function $\mathcal{T} : \mathcal{Y} \times \mathcal{X} \to \mathcal{Y}$ which satisfies the following properties:

1) For every $x \in \mathcal{X}$, the function $\mathcal{T}(\cdot, x) : \mathcal{Y} \to \mathcal{Y}$ is bijective.
2) For every $x_1, x_2 \in \mathcal{X}$, the following two relations hold:

$$p(y|x_1) = p(\mathcal{T}(y, x_2 - x_1)|x_2) \tag{1}$$

and

$$\mathcal{T}(\mathcal{T}(y, x_1), x_2) = \mathcal{T}(y, x_1 + x_2). \tag{2}$$

When dealing with channels whose output alphabet is continuous, an additional requirement on the mapping $\mathcal{T}$ is that its Jacobian equals to one.[1] In this case, the condition in (1) implies that

$$\int p(y|x_1)\, dy = \int p(\mathcal{T}(y, x_2 - x_1)|x_2)\, dy.$$

For the particular case of channels with a binary-input alphabet, and whose output alphabet $\mathcal{Y}$ is the set of real numbers, setting

$$\mathcal{T}(y, x) = \begin{cases} y & x = 0 \\ -y & x = 1 \end{cases}$$

the definition coincides with the well known output-symmetric definition for MBIOS channels.

In some cases, block codes are applied to arbitrary channels together with a random coset mechanism (see for example [11], [13], [14]). That is, instead of transmitting the coded message $\mathbf{x}$, the vector $\mathbf{x}+\mathbf{v}$ is transmitted where $\mathbf{v}$ is a random vector, called the coset, known to both the transmitter and the receiver, and the addition is carried out symbol-wise. When coding schemes with a random coset mechanism are applied to an arbitrary memoryless channel, the symmetry (according to Definition 1) of the equivalent channel is guarantied. To see this, consider the equivalent channel that includes the addition of the coset symbols followed by the original channel, and whose observations are pairs (y, v), where v is the random coset symbol added to the transmitted coded symbol, and y is the (original) channel output. Assuming a memoryless channel, the symmetry is guarantied by setting

$$\mathcal{T}((y, v), x) = (y, v - x), \quad y \in \mathcal{Y},\ x, v \in \mathcal{X}$$

where $\mathcal{X}$ and $\mathcal{Y}$ are the input and output alphabets, respectively. Notice that $\mathcal{T}$ is now defined over $(\mathcal{Y} \times \mathcal{X}) \times \mathcal{X}$, where $\mathcal{Y} \times \mathcal{X}$ forms the output alphabet of the equivalent channel.

For MBIOS channels, the capacity is attained with a uniform input distribution. In addition, random coding with a uniform (and memoryless) distribution attains the optimum random-coding error exponent provided by Gallager [2], [13], [33]. The following Lemma generalizes these results for the case of discrete, memoryless, and symmetric channels according to Definition 1 (a similar result follows for the case of memoryless symmetric channels with a continuous output-alphabets). The lemma is also valid for symmetric DMC's in the sense defined by Gallager in [13, p. 94].

[1] It is possible to use a generalized definition for both discrete and continuous output alphabets using the notion of unitary functions as done for example in [21].

Lemma 1. Let Q be a probability function over the input alphabet $\mathcal{X}$, and let p be a transition probability function of a discrete symmetric and memoryless channel. Then, the mutual information $I(Q)$, between the channel input (with an input probability distribution Q) and the channel output, given by

$$I(Q) = \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}} Q(x)p(y|x) \ln\left(\frac{p(y|x)}{\sum_{x'\in\mathcal{X}} Q(x')p(y|x')}\right)$$

and the Gallager function $E_0(\rho, Q)$ [13], defined by

$$E_0(\rho, Q) \triangleq -\ln\left(\sum_{y\in\mathcal{Y}}\left(\sum_{x\in\mathcal{X}} Q(x)p(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}\right), \ \rho \geq 0$$

are maximized (for every $\rho \geq 0$) by a uniform distribution.

Proof: The proof follows trivially by applying [33, Theorems 3.2.2 & 3.2.3] to the case at hand. ■

Consider linear block codes over the non-binary alphabet $\mathcal{X}$. Specifically, let $\mathbf{G}$ be a $k\times n$ matrix with components over the alphabet $\mathcal{X}$. Then, the linear block code with a generator matrix $\mathbf{G}$, denoted by $\mathcal{C} = \{\mathbf{x}_m\}_{m=1}^{q^k}$ where $\mathbf{x}_m = (x_{m,1}, \ldots, x_{m,n})$, is the set of all q^k linear combinations of the rows of $\mathbf{G}$. The conditional error probability of the m-th message is given according to

$$P_{\mathrm{e}|m} \ = \ \sum_{\mathbf{y}\in\Lambda_m^{\mathrm{c}}} p(\mathbf{y}|\mathbf{x}_m)$$

where Λ_m forms the decision region for the m-th codeword, and the superscript 'c' stands for the complementary set. The decision region of the m-th codeword under ML decoding gets the form

$$\Lambda_m = \left\{\mathbf{y} : p(\mathbf{y}|\mathbf{x}_m) > p(\mathbf{y}|\mathbf{x}_{m'}), \ \forall\, m' \neq m\right\}$$

and ties are resolved randomly with equal probability. A well-known result for binary linear block codes operating over MBIOS channels is that their error probability under ML decoding is independent of the actual transmitted codeword. This result enables a great simplification to the error performance analysis by assuming that the all-zero codeword, designated by $\mathbf{0}$, is transmitted. The following proposition is a generalization of this result for linear block codes communicated over memoryless and symmetric channels whose input alphabet is discrete.

Proposition 1. Let $\mathcal{C}$ be a linear block code whose transmission takes place over a memoryless and symmetric channel. Then, the block error probability under ML decoding is independent of the transmitted codeword.

Proof: See Appendix A. ■

Note that in contrast to Lemma 1, Proposition 1 does not necessarily hold for symmetric DMCs in the broader sense, as defined by Gallager in [13, p. 94]. This is demonstrated in the following counter-example.

Example 1. Consider a DMC with the integer ring $\mathbb{Z}_4$ (with arithmetic operations modulo-4) as a common input and output alphabets, and with the following transition probability matrix.

$$\begin{pmatrix} 0.20 & 0.24 & 0.30 & 0.26 \\ 0.30 & 0.20 & 0.26 & 0.24 \\ 0.24 & 0.26 & 0.20 & 0.30 \\ 0.26 & 0.30 & 0.24 & 0.20 \end{pmatrix}.$$

In this matrix, the element at the i–th row and j–th column (where $i, j \in \{1, \ldots, 4\}$) refers to the transition probability when the channel input is equal to $i-1$ and the corresponding channel output is equal to $j-1$. Although it is not a symmetric matrix, this transition probability matrix is symmetric in the sense defined by Gallager (notice that each row and column is a permutation of another row and column, respectively). However, when applying the linear block code $\{00, 13, 22, 31\}$ to the considered channel, the resulting conditional error probabilities under ML decoding are 0.7540, 0.7210, 0.5424 and 0.7210, respectively, and they therefore depend on the transmitted codeword. To show this, we first need to determine the ML decoding regions for the considered

code and channel. This is accomplished by evaluating the conditional probabilities of each possible output pair given each possible transmitted codeword (e.g., $p(03|31) = 0.26 \cdot 0.24 = 0.0624$). The decoding region for the all-zero codeword 00 is the set $\{22, 23, 32\}$ (note that the '00' vector is not included in the decision region of this codeword, and on the other hand, the vector '22' which forms a codeword is included in the decision region of the all-zero codeword). The conditional error probability given that the all-zero codeword is transmitted is therefore equal to $1 - p(22|00) - p(23|00) - p(32|00) = 1 - 0.30^2 - 0.30 \cdot 0.26 - 0.26 \cdot 0.30 = 0.7540$. The rest of the conditional error probabilities are similarly evaluated.

III. Gallager-type Bounds for Memoryless Symmetric Channels

A. The DS2 bound

Let $\mathcal{C}$ be an (n, k) linear block code defined over the input-alphabet $\mathcal{X}$ with a cardinality q. Consider the conditional error probability under ML decoding given that the m^{th} message is transmitted, denoted by $P_{\text{e}|m}$. The DS2 bound [3], [4] on this conditional error probability gets the form

$$P_{\text{e}|m} \leq \left(\sum_{\mathbf{y} \in \mathcal{Y}^n} G_n^m(\mathbf{y}) p_n(\mathbf{y}|\mathbf{x}_m) \right)^{1-\rho} \cdot \left\{ \sum_{m' \neq m} \sum_{\mathbf{y} \in \mathcal{Y}^n} G_n^m(\mathbf{y})^{1-\frac{1}{\rho}} p_n(\mathbf{y}|\mathbf{x}_m) \left(\frac{p_n(\mathbf{y}|\mathbf{x}_{m'})}{p_n(\mathbf{y}|\mathbf{x}_m)} \right)^{\lambda} \right\}^{\rho} \tag{3}$$

where $\mathcal{Y}$ is a discrete output-alphabet, $G_n^m(\mathbf{y})$ is an arbitrary non-negative function of $\mathbf{y} \in \mathcal{Y}^n$, and $0 \leq \rho \leq 1$ and $\lambda \geq 0$ are arbitrary real-valued parameters (see [1], [5], [6], [34]). Here $p_n(\mathbf{y}|\mathbf{x})$ designates the transition probability of the channel where $\mathbf{x} \in \mathcal{C}$ is a transmitted codeword and $\mathbf{y} \in \mathcal{Y}^n$ is the received vector. Notice that this bound holds for an arbitrary channel regardless of its input alphabet (binary or non-binary).

Consider now the class of memoryless symmetric channels with an input-alphabet $\mathcal{X}$. According to Proposition 1, $P_{\text{e}|m}$ is independent of the transmitted message m. We further assume that $G_n^0(\mathbf{y})$ is expressed in the following product form:

$$G_n^0(\mathbf{y}) = \prod_{i=1}^{n} g(y_i).$$

The following bound on the decoding error probability is obtained for a discrete output alphabet (a similar proposition can be stated for channels with continuous output alphabet):

Proposition 2. Consider an (n, k) linear block code $\mathcal{C}$ whose transmission takes place over a memoryless symmetric channel. Assume that the channel input and output alphabets are $\mathcal{X}$ and $\mathcal{Y}$, respectively, and let p be the transition probability of the channel. Then the block error probability of the code $\mathcal{C}$ under ML decoding, P_{e}, satisfies

$$P_{\text{e}} \leq \left(\sum_{y \in \mathcal{Y}} g(y) p(y|0) \right)^{n(1-\rho)} \left\{ \sum_{m' \neq 0} \prod_{i=1}^{n} \sum_{y \in \mathcal{Y}} g(y)^{1-\frac{1}{\rho}} p(y|0) \left(\frac{p(y|x_{m',i})}{p(y|0)} \right)^{\lambda} \right\}^{\rho} \tag{4}$$

where $g : \mathcal{Y} \to \mathbb{R}$ is an arbitrary non-negative real function, $\lambda \geq 0$, and $0 \leq \rho \leq 1$ are arbitrary real-valued parameters.

Proof: See Appendix B. ∎

B. Performance evaluation of ensembles of linear block codes

The following technical lemma considers the error probability under ML decoding of an ensemble of linear block codes.

Lemma 2. Let $\mathcal{C}$ be an ensemble of linear block codes with a block length n, and let $d_{\min}$ be the random variable designating the minimum Hamming distance of a randomly selected codebook from this ensemble. Assume that there exist non-negative numbers D_n and ϵ_n, such that

$$\sum_{\{\mathbf{t}\in\mathcal{T}^*:\ n-t_0\leq D_n\}} \mathsf{E}\big[|\mathcal{C}_{\mathbf{t}}|\big] \leq \epsilon_n \tag{5}$$

where $\mathsf{E}\big[\,|\mathcal{C}_{\mathbf{t}}|\,\big]$ denotes the expected number of codewords with a composition $\mathbf{t}$, and $\mathcal{T}^*$ denotes the entire set of compositions except for the one of the all-zero codeword. Then, the block error probability under ML decoding satisfies

$$P_{\mathrm{e}} \leq \Pr(\text{ error } \mid d_{\min} > D_n) + \epsilon_n. \tag{6}$$

Proof:

$$\begin{aligned} P_{\mathrm{e}} &= \Pr(\text{ error } \mid d_{\min} > D_n)\Pr(d_{\min} > D_n) \\ &\quad + \Pr(\text{ error } \mid d_{\min} \leq D_n)\Pr(d_{\min} \leq D_n) \\ &\leq \Pr(\text{ error } \mid d_{\min} > D_n) + \Pr(d_{\min} \leq D_n). \end{aligned}$$

Let C be a codebook, chosen uniformly at random from the code ensemble $\mathcal{C}$, and let $w_{\mathrm{H}}(\mathbf{c})$ denote the Hamming weight of a codeword $\mathbf{c}$. Then, (6) follows from (5) since the union bound gives that

$$\begin{aligned} \Pr(d_{\min} \leq D_n) &\leq \sum_{\{\mathbf{c}\neq\mathbf{0}:\ w_{\mathrm{H}}(\mathbf{c})\leq D_n\}} \Pr(\mathbf{c}\in C) \\ &= \sum_{\{\mathbf{t}\in\mathcal{T}^*:\ n-t_0\leq D_n\}} \mathsf{E}\big[|\mathcal{C}_{\mathbf{t}}|\big]. \end{aligned} \tag{7}$$

■

The following theorem provides an upper bound on the decoding error probability of ensembles of linear block codes whose transmission takes place over memoryless symmetric channels.

Theorem 1. Under the assumptions and notation in Proposition 2 and Lemma 2, the block error probability under ML decoding satisfies

$$\begin{aligned} P_{\mathrm{e}} \leq &\left(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\right)^{n(1-\rho)} \\ &\cdot\left(\sum_{\mathbf{t}\in\mathcal{T}^*:\ n-t_0>D_n} \mathsf{E}\Big[|\mathcal{C}_{\mathbf{t}}| \ \mid d_{\min} > D_n\Big] \prod_{x\in\mathcal{X}} \Big(\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}} p(y|0)^{1-\lambda} p(y|x)^{\lambda}\Big)^{t_x}\right)^{\rho} + \epsilon_n \end{aligned} \tag{8}$$

where $\mathsf{E}\big[\,|\mathcal{C}_{\mathbf{t}}| \ \mid d_{\min} > D_n\big]$ denotes the expected number of codewords whose composition is equal to $\mathbf{t}$, under the requirement that the minimal Hamming weight of the randomly selected codebook is larger than D_n (i.e., $d_{\min} > D_n$).

Proof: Following the notation in Proposition 2, let $0 \leq \rho \leq 1$, $\lambda \geq 0$, and $g : \mathcal{Y} \to \mathbb{R}$ be an arbitrary non-negative function. Applying Proposition 2 for upper bounding the first summand in (6), it follows from (4) that

$$\begin{aligned} &\Pr(\text{ error } \mid d_{\min} > D_n) \leq \\ &\qquad \left(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\right)^{n(1-\rho)} \mathsf{E}\left[\left(\sum_{\mathbf{t}\in\mathcal{T}^*}\sum_{\mathbf{c}\in\mathcal{C}_{\mathbf{t}}}\prod_{i=1}^{n}\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}} p(y|0)\left(\frac{p(y|c_i)}{p(y|0)}\right)^{\lambda}\right)^{\rho} \ \middle|\ d_{\min} > D_n\right] \end{aligned}$$

where $\mathcal{C}_{\mathbf{t}}$ is the set of all codewords whose composition is $\mathbf{t}$, in an arbitrary codebook; the statistical expectation is taken over all codebooks whose Hamming distance is larger than D_n, where these codebooks are selected with a uniform probability. For an arbitrary codebook, notice that the double summations in the RHS of the last inequality,

over compositions $\mathbf{t}$ and codewords $\mathbf{c} \in \mathcal{C}_{\mathbf{t}}$, is equivalent to a single summation over all non-zero codewords. Using Jensen's inequality $\mathsf{E}[X^\rho] \leq \big(\mathsf{E}[X]\big)^\rho$ which holds for $0 \leq \rho \leq 1$, since the channel is memoryless we get

$$\begin{aligned}
&\Pr(\text{ error } \mid d_{\min} > D_n) \\
&\quad \leq \quad \left(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\right)^{n(1-\rho)} \\
&\qquad \cdot \left(\sum_{\mathbf{t}\in\mathcal{T}^*} \mathsf{E}\left[\sum_{\mathbf{c}\in\mathcal{C}_{\mathbf{t}}} \prod_{x\in\mathcal{X}} \left(\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}} p(y|0) \left(\frac{p(y|x)}{p(y|0)}\right)^{\lambda}\right)^{t_x} \;\middle|\; d_{\min} > D_n\right]\right)^{\rho} \\
&\quad = \quad \left(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\right)^{n(1-\rho)} \\
&\qquad \cdot \left(\sum_{\mathbf{t}\in\mathcal{T}^*} \mathsf{E}\left[|\mathcal{C}_{\mathbf{t}}| \;\middle|\; d_{\min} > D_n\right] \prod_{x\in\mathcal{X}} \left(\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}} p(y|0) \left(\frac{p(y|x)}{p(y|0)}\right)^{\lambda}\right)^{t_x}\right)^{\rho}. \qquad (9)
\end{aligned}$$

Next, for all codewords whose composition $\mathbf{t}$ satisfies $n - t_0 \leq D_n$, their Hamming weight is not larger than D_n. As a result

$$\mathsf{E}\left[|\mathcal{C}_{\mathbf{t}}| \;\mid d_{\min} > D_n\right] = 0, \quad \forall\ \mathbf{t} \in \mathcal{T} \text{ where } n - t_0 \leq D_n \qquad (10)$$

and (8) follows from Lemma 2, (9), and (10). ■

The following theorem is a particularization of Theorem 1:

Theorem 2. Under the assumptions and notation in Theorem 1, the block error probability satisfies

$$P_{\mathrm{e}} \leq \; q^{-nE_{\mathrm{r}}\left(R+\frac{\log_q \alpha(\mathcal{C},D_n)}{n}\right)} \; + \; \epsilon_n \qquad (11)$$

where n and R are the block length and code rate (measured in q-ary symbols per channel use), respectively, and

$$\begin{aligned}
E_{\mathrm{r}}(R) &\triangleq \max_{0\leq\rho\leq 1} \left(E_0(\rho) - \rho R\right) \\
E_0(\rho) &\triangleq -\log_q \left\{\sum_{y\in\mathcal{Y}} \left[\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right]^{1+\rho}\right\} \\
\alpha(\mathcal{C}, D_n) &\triangleq \max_{\{\mathbf{t}\in\mathcal{T}^*:\ n-t_0>D_n\}} \left\{\frac{\mathsf{E}\left[|\mathcal{C}_{\mathbf{t}}| \;\middle|\; d_{\min} > D_n\right]}{q^{-n(1-R)}\binom{n}{\mathbf{t}}}\right\}. \qquad (12)
\end{aligned}$$

Proof: It follows from (8) that

$$\begin{aligned}
&\Pr(\text{ error } \mid d_{\min} > D_n) \\
&\leq \left(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\right)^{n(1-\rho)} q^{-n\rho(1-R)} \\
&\quad \cdot \left(\sum_{\mathbf{t}\in\mathcal{T}^*:\ n-t_0>D_n} \frac{\mathsf{E}\left[|\mathcal{C}_{\mathbf{t}}| \ \middle|\ d_{\min} > D_n\right]}{q^{-n(1-R)}\binom{n}{\mathbf{t}}} \binom{n}{\mathbf{t}} \prod_{x\in\mathcal{X}} \left(\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}} p(y|0)^{1-\lambda} p(y|x)^{\lambda}\right)^{t_x}\right)^{\rho} \\
&\leq \left(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\right)^{n(1-\rho)} q^{-n\rho(1-R)} \cdot \left(\max_{\mathbf{t}\in\mathcal{T}^*:\ n-t_0>D_n} \left\{\frac{\mathsf{E}\left[|\mathcal{C}_{\mathbf{t}}| \ \middle|\ d_{\min} > D_n\right]}{q^{-n(1-R)}\binom{n}{\mathbf{t}}}\right\}\right)^{\rho} \\
&\quad \cdot \left\{\sum_{\mathbf{t}\in\mathcal{T}^*:\ n-t_0>D_n} \binom{n}{\mathbf{t}} \prod_{x\in\mathcal{X}} \left(\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}} p(y|0)^{1-\lambda} p(y|x)^{\lambda}\right)^{t_x}\right\}^{\rho} \\
&\overset{(a)}{=} q^{-n\rho(1-R)} \left(\alpha(\mathcal{C}, D_n)\right)^{\rho} \left(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\right)^{n(1-\rho)} \left[\sum_{l=D_n+1}^{n} \binom{n}{l} \left(\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}} p(y|0)\right)^{n-l} \right. \\
&\qquad \left. \sum_{\substack{t_1,\ldots,t_{q-1} \\ t_1+\ldots+t_{q-1}=l}} \binom{l}{t_1,\ldots,t_{q-1}} \prod_{x\in\mathcal{X}^*} \left(\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}} p(y|0)^{1-\lambda} p(y|x)^{\lambda}\right)^{t_x}\right]^{\rho} \\
&= q^{-n\rho(1-R)} \left(\alpha(\mathcal{C}, D_n)\right)^{\rho} \left(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\right)^{n(1-\rho)} \\
&\quad \cdot \left[\sum_{l=D_n+1}^{n} \binom{n}{l} \left(\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}} p\left(y|0\right)\right)^{n-l} \left(\sum_{x\in\mathcal{X}^*} \sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}} p(y|0)^{1-\lambda} p(y|x)^{\lambda}\right)^{l}\right]^{\rho} \qquad (13)
\end{aligned}$$

where the equality in (a) is due to (12). Setting

$$g(y) = \left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{\rho} p(y|0)^{-\frac{\rho}{1+\rho}}, \quad \lambda = \frac{1}{1+\rho} \qquad (14)$$

and using the symmetry of the channel, it is proved in Appendix C that the upper bound in (13) is transformed to

$$\Pr(\text{ error } \mid d_{\min} > D_n) \leq q^{-nE_{\mathrm{r}}\left(R+\frac{\log_q \alpha(\mathcal{C},D_n)}{n}\right)}. \qquad (15)$$

Finally, the proof follows from Lemma 2 and (15). ■

A similar theorem can be stated for memoryless symmetric channels with continuous-output alphabets, where sums are replaced by integrals (some improvement can be achieved by applying a similar modification to the one provided in [35] for the class of MBIOS channels).

The bound in Theorem 2 is based on two summands. The first is an adaptation of the SFB to non-binary linear block codes which applies to the codebooks whose minimum distance exceeds an arbitrary threshold D_n. The second term relates to the probability that a randomly selected codebook from the ensemble has a minimum Hamming distance which does not exceed D_n. As a result, the second term on the RHS of (11) does not depend on the communication channel, but only depends on the code ensemble and the arbitrary threshold D_n. This is in contrast to [11] and [36] where no such separation of codebooks is used. The SFB in [11], [36] is combined with a union bound corresponding to all pairwise error probabilities of relevant codewords and it depends on the

communication channel (via the Bhattacharyya parameter of the channel). In addition, the derivation of the SFB part itself for memoryless symmetric channels in Theorem 2, completely differs from the one applied in [11] to the case of coding with a random coset mechanism. Note that the random coding mechanism for an arbitrary memoryless channel forms a memoryless and symmetric channel according to Definition 1. As a result, the SFB in [11] follows as a particular case of Theorem 2 (the same goes for [15] where the considered modulo-additive noise channel is also symmetric according to Definition 1).

In general, the conditional expectation of the composition spectrum given that the minimum Hamming distance exceeds a certain positive threshold D_n (i.e., $\mathsf{E}\big[|\mathcal{C}_{\mathbf{t}}|\big|d_{\min} > D_n\big]$) is not available. Nevertheless, it is possible to use the inequality

$$\mathsf{E}\Big[|\mathcal{C}_{\mathbf{t}}|\Big] \geq \mathsf{E}\Big[|\mathcal{C}_{\mathbf{t}}| \mid d_{\min} > D_n\Big] \ (1-\epsilon_n) \tag{16}$$

where the RHS of this inequality requires the knowledge of the expectation of the complete composition spectrum $\mathsf{E}\big[|\mathcal{C}_{\mathbf{t}}|\big]$. The inequality in (16) follows from (5) and (7) since

$$\begin{aligned}
\mathsf{E}\Big[|\mathcal{C}_{\mathbf{t}}|\Big] &= \mathsf{E}\Big[|\mathcal{C}_{\mathbf{t}}| \mid d_{\min} > D_n\Big] \ \Pr(d_{\min} > D_n) \\
&\quad + \mathsf{E}\Big[|\mathcal{C}_{\mathbf{t}}| \mid d_{\min} \leq D_n\Big] \ \Pr(d_{\min} \leq D_n) \\
&\geq \mathsf{E}\Big[|\mathcal{C}_{\mathbf{t}}| \mid d_{\min} > D_n\Big] \ \Pr(d_{\min} > D_n) \\
&\geq \mathsf{E}\Big[|\mathcal{C}_{\mathbf{t}}| \mid d_{\min} > D_n\Big] \ (1-\epsilon_n).
\end{aligned} \tag{17}$$

Applying this to the RHS of the bound in Theorem 1, gives a looser version of this bound but is more amenable for analysis. The same inequality is valid when expurgation of codebooks is considered. The expurgated ensemble is constructed by removing from the original ensemble, all codebooks whose minimum Hamming distance is not above D_n. Since all codebooks in the expurgated ensemble have a minimum distance greater than D_n, then the additive term ϵ_n on the RHS of (8) vanishes.

Consider an ensemble of linear block codes and pick a codebook from this ensemble uniformly at random. We consider in the following ensembles having the property that the probability that a vector is a codeword only depends on its Hamming weight (so all vectors of a fixed composition are codewords with equal probability). As a result, the expected complete composition spectrum $\mathsf{E}\,|\mathcal{C}_{\mathbf{t}}|$ satisfies

$$\mathsf{E}\Big[|\mathcal{C}_{\mathbf{t}}|\Big] = P(n-t_0)\binom{n}{\mathbf{t}} \tag{18}$$

where $P(l)$ denotes the probability that a word whose Hamming weight is l, forms a codeword in a randomly selected codebook from the ensemble. Assuming (18), the evaluation of $\alpha\,(\mathcal{C})$ in Theorem 2 is considerably reduced. In addition, it is possible to provide a refinement of the analysis used for the derivation of Theorem 2. Specifically, we tighten the bound by circumventing the need to take the maximization involved in the evaluation of $\alpha\,(\mathcal{C})$ in (12), and still obtain a bound which is computationally feasible due to the assumption in (18). This results in the following theorem:

Theorem 3. Under the assumptions and notation in Theorem 1, for ensembles satisfying (18), the block error probability satisfies

$$P_{\mathbf{e}} \leq A(\rho)^{n(1-\rho)} \left(\sum_{D_n < l \leq n} \left(\frac{P(l)}{1-\epsilon_n} \binom{n}{l} B(\rho)^{n-l} C(\rho)^l \right) \right)^{\rho} + \epsilon_n \tag{19}$$

where $0 \leq \rho \leq 1$, ϵ_n is defined in (6), and

$$\begin{aligned}
A(\rho) &\triangleq \sum_{y\in\mathcal{Y}} \left(\frac{1}{q} \sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}} \right)^{1+\rho} \\
B(\rho) &\triangleq \sum_{y\in\mathcal{Y}} \left(\frac{1}{q} \sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}} \right)^{\rho-1} \left(\frac{1}{q} \sum_{x\in\mathcal{X}} p(y|x)^{\frac{2}{1+\rho}} \right) \\
C(\rho) &\triangleq qA(\rho) - B(\rho).
\end{aligned}$$

Proof: See Appendix D. ■

C. Performance of non-binary regular LDPC ensembles

The non-binary (c, d)-regular LDPC ensemble, proposed by Gallager in [22, Ch. 5], is considered with the q-ary symmetric channel and the AWGN channel with a q-ary PSK modulation (both channels are symmetric according to Definition 1). The Gallager ensemble is defined using a sparse parity-check matrix with binary elements. This matrix is regular, having c ones in each column and d ones in each row. The LDPC ensemble is constructed as follows:

1) Divide the parity check matrix into c sub-matrices.
2) Fill the first sub-matrix with ones in a descending order.
3) All other sub-matrices are chosen as random permutations of the first sub-matrix.
4) Parity-check equations are evaluated using a modulo-q arithmetics.

The following lemma is provided in [22] which implies an upper bound on the complete composition spectrum satisfying the condition in (18):

Lemma 3. Consider the regular non-binary LDPC ensemble of Gallager. Let $\mathbf{x}$ be a vector of weight $l > 0$. The probability $P(l)$ that the vector $\mathbf{x}$ is a codeword of a codebook which is selected uniformly at random from the ensemble, is upper bounded by

$$P(l) \leq \left(\frac{\exp\left(\frac{n}{d}\left(\mu_q(s) - s\mu_q'(s) + (d-1)\ln q\right)\right)}{\binom{n}{l}(q-1)^l} \right)^c \tag{20}$$

where

$$\mu_q(s) \triangleq \ln\left(\frac{\left(1+(q-1)e^s\right)^d + (q-1)\left(1-e^s\right)^d}{q^d} \right)$$

and s is a real number given by the solution of the following equation

$$\frac{n}{d}\mu_q'(s) = l. \tag{21}$$

Note, that the bound in (20) is valid for all s, not only for the one satisfying (21) which yields the minimum bound in (20). Using the change of variables $s = \ln\frac{1-u}{1+(q-1)u}$, $-\frac{1}{q-1} \leq u \leq 1$, in (21), results in the following polynomial equation:

$$\left(\frac{wq}{n} - 1\right)u^d + u^{d-1} + u + \frac{wq}{n(q-1)} - 1 = 0.$$

For $q > 2$, this equation has a single root in the interval $\left[-\frac{1}{q-1}, 1\right]$ (the details concerning the evaluation of RHS of (20) in the binary case are provided in [8]).

The 1961 Gallager-Fano bound (see [1], [22]) and Lemma 3 imply an exponential decrease (in terms of the block length) of the decoding error probability for the considered LDPC ensemble with an expurgation which dismiss all the codebook with minimal Hamming distance which is below a certain threshold that is scaling linearly with the block length (this result is elaborated for the binary case in [36]). In addition, in a sharp contrast to the symmetry in Definition 1, the bounding in [22] requires an implicit definition of symmetry which is tailored to the specific considered bound, that is the 1961 Gallger-Fano bound (see [22, p. 51]). Nevertheless, it is indicated by Divsalar [6] that the DS2 bound, applied in the following, is superior over the 1961 Gallager-Fano bound when applied to a particular code or a structured ensemble.

Following [37], the performance of the binary regular LDPC ensemble of Gallager [2] is evaluated in [32] via the tangential sphere bound (see [1], [38]) where the upper bound on the distance spectrum provided in [2] and applied in [37] is replaced in [32] by the exact distance spectrum. The following lemma generalizes the ideas in [32] for the non-binary LDPC ensemble:

Lemma 4. Under the assumptions and notation in Lemma 3, the probability $P(l)$ satisfies

$$P(l) = \left(\frac{A_l}{\binom{n}{l}(q-1)^l} \right)^c, \quad 2 \leq l \leq n \tag{22}$$

where

$$\sum_{2\leq l\leq n} A_l X^l \;=\; \big(A^*(X)\big)^{\frac{n}{d}} \tag{23}$$

$$A^*(X) \;\triangleq\; 1+\sum_{l=2}^{d}\binom{d}{l}a_0(l)X^l \tag{24}$$

$$\begin{pmatrix} a_0(l)\\ a_1(l)\\ \vdots\\ \\ a_{q-1}(l)\end{pmatrix} = \begin{pmatrix} 0 & 1 & \cdots & 1 & 1\\ 1 & 0 & 1 & \cdots & 1\\ \vdots & & \ddots & & \\ 1 & \cdots & 1 & 0 & 1\\ 1 & & \cdots & 1 & 0\end{pmatrix}^{l-1}_{q\times q}\begin{pmatrix} 0\\ 1\\ \vdots\\ \\ 1\end{pmatrix}_{q\times 1}. \tag{25}$$

Proof: Denote by $a_{x^*}(l)$, $1\leq l\leq d$, the number of choices of l non-zero elements in $\{1,\ldots,q-1\}$ whose summation modulo q equals x^*, $0\leq x^*\leq q-1$. Then, there are $\binom{d}{l}a_{x^*}(l)$ vectors $\mathbf{x}=(x_1,\ldots,x_d)$, whose Hamming weight is l, satisfying

$$x_1+\cdots+x_d = x^* \mod q.$$

As a result,$A^*(X)$ in (24) is the enumerator for the number of vectors $\mathbf{x}$ satisfying the parity-check equation

$$x_1+\cdots+x_d = 0 \mod q$$

and the enumerator of the first sub-matrix in the considered ensemble is given in (23) (this resembles the idea provided in [32] for the binary case). Next, notice that the sequences $a_{x^*}(l)$ satisfy the following system of recursive equations:

$$a_{x^*}(l)=\sum_{x=1}^{q-1} a_{(x^*-x)\bmod q}(l-1),\ 0\leq x^*\leq q-1$$

with the initial conditions $a_0\,(1)=0$, and $a_x\,(1)=1$ for all $x\neq 0$. Using a vector notation, the equations are written according to

$$\begin{pmatrix} a_0(l)\\ a_1(l)\\ \vdots\\ \\ a_{q-1}(l)\end{pmatrix} = \begin{pmatrix} 0 & 1 & \cdots & 1 & 1\\ 1 & 0 & 1 & \cdots & 1\\ \vdots & & \ddots & & \\ 1 & \cdots & 1 & 0 & 1\\ 1 & & \cdots & 1 & 0\end{pmatrix}_{q\times q}\begin{pmatrix} a_0(l-1)\\ a_1(l-1)\\ \vdots\\ \\ a_{q-1}(l-1)\end{pmatrix}$$

whose solution is given in (25). Finally, (22) is established in [22] which concludes the proof. ■

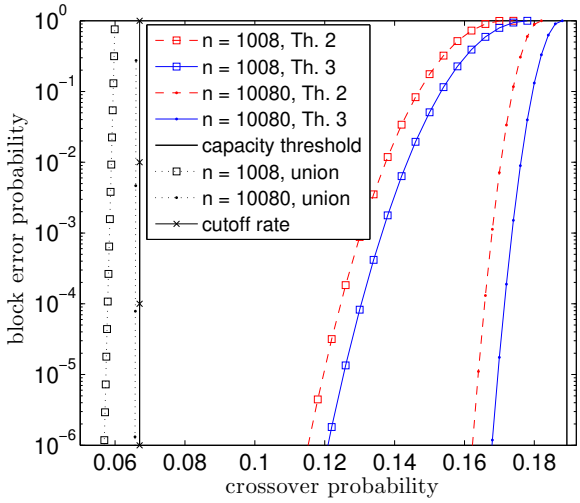
Example 2 (q-ary symmetric channels)**.** The block error probabilities for the considered LDPC ensemble under ML decoding are presented in Figure 1 when the transmission takes place over the q-ary symmetric channel (where $q=4,8$). The performance bounds introduced in this paper are compared to the union bound in order to show that the latter bound is useless beyond the crossover probability which corresponds to the cutoff rate. More specifically, for a q-ary symmetric channel, the cutoff rate is given by

$$R_0 = 1-2\log_q\left(\sqrt{1-p}+\sqrt{p(q-1)}\right)$$

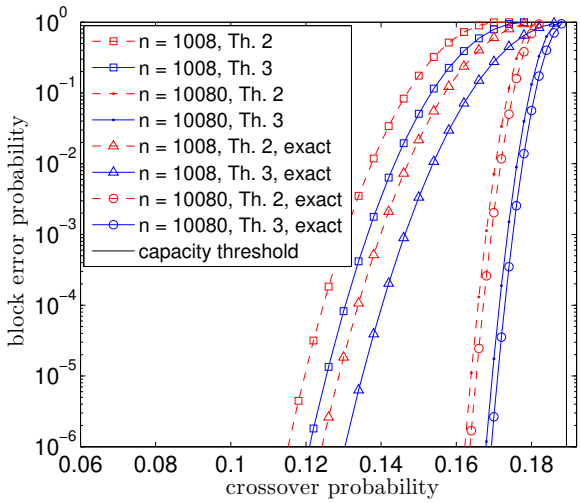
so the crossover probability which follows by setting the value of R_0 to the code rate (which is one-half symbol per channel use in Fig. 1) is equal to $p=0.0670$ and $p=0.0739$ for quaternary and octal input alphabets, respectively. The union bound shown in the upper plot of Fig. 1 (see plot (a)) has a sharp decline around the crossover probability which corresponds to the cutoff rate of the q-ary symmetric channel (i.e., around $p=0.0670$ for $q=4$). Plot (a) also exemplifies the potential application of the proposed bounds in this paper to assess properly the performance of efficient code ensembles (which perform reliably at rates exceeding the cutoff rate of the channel). Fig. 1 provides a zoom in of plot (a) where the focus is on the improved bounds provided in Theorems 2 and 3. The performance is evaluated in Figure 1(b) using Theorem 2 (in dashed lines) and Theorem 3 (in solid lines), applied to the $(8,16)$

Gallager expurgated ensemble with a quaternary alphabet. The ensemble spectrum is upper bounded via Lemma 3 and in addition it is exactly evaluated using Lemma 4; both options are applied in this example so that the possible improvement is demonstrated. As suggested in [32], the numerical evaluation of the exponent in (23) is carried out via the binary method (see [39, p. 441]). The decoding error probability for a block length of $n = 1008$ symbols is shown with a square-marker where the bound in (19) is evaluated with $D_n = 99$ and a corresponding $\epsilon_n = 1.2 \cdot 10^{-4}$ when the ensemble spectrum is upper bounded using Lemma 3. When the exact spectrum is evaluated using Lemma 4, the decoding error probability for the same block length is shown with a triangle-marker (and its corresponding ϵ_n value is around 10^{-11}). The SFB in (11) is applicable only for higher values of D_n, and is presented for $D_n = 173$ (and a corresponding $\epsilon_n = 0.1$ when Lemma 3 is used, and around 10^{-11} when Lemma 4 is used). For a block length of $10,080$ symbols (presented with a dot-marker and a circle-marker, when Lemma 3 and Lemma 4 are used, respectively), the bound provided by Theorem 3 is presented for $D_n = 600$ (and a corresponding $\epsilon_n = 10^{-7}$ and 10^{-17}, when Lemma 3 and Lemma 4 are used, respectively). The SFB, is applicable only for higher values of D_n (the presented curve is for $D_n = 1834$ symbol, with a corresponding $\epsilon_n = 0.11$ and 10^{-17}, respective to the Lemma 3 and Lemma 4). It is evident that the bound in Theorem 2 is looser as compared to the bound in Theorem 3. In addition, the bound in (19) maintains its performance for a considerable range of D_n values. The considered bounds are further improved when the upper bound for the weight spectrum in Lemma 3 is replaced with the one in Lemma 4. The inferiority of the SFB in (11) is further pronounced for higher alphabet sizes, as demonstrated for octal signaling: The bound in (19) is shown in Figure 1(c) with D_n values of 119 and 887, respective to $n = 1008$ and $n = 10080$ symbols. The bound in (11), which requires higher values of D_n, is shown for the D_n values of 191 and 1951, respectively. In both cases, the ϵ_n values are 10^{-5}, and 10^{-9}, respectively, when the upper bound in Lemma 3 is used, so a very small fraction of the codebooks is actually expurgated.

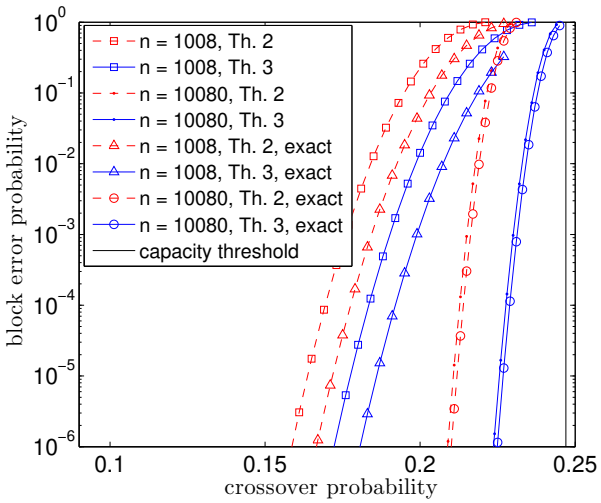
Example 3 (AWGN channels with a q-ary PSK modulation)**.** The block error probability under ML decoding for the considered ensemble with an alphabet size of $q = 4, 8, 16$, and 32 is depicted in Figure 2 when the transmission takes place over the AWGN channel with a q-ary PSK modulation. All presented bounds are evaluated with the exact composition spectrum in Lemma 4. The bound in Theorem 3 is evaluated for an alphabet size of $q = 4$ with D_n values of 38 and 282, respective to the presented block lengths (i.e., $n = 1008$ and 10080). For $q = 8$, the two respective curves refer to D_n values of 23 and 216; for $q = 16$, they refer to D_n values of 15 and 132, and for $q = 32$, they refer to D_n values of 12 and 102. The bound provided in Theorem 3 maintains its tightness for a considerably large range of D_n values and alphabet cardinalities; in fact this bound is informative up to the ultimate threshold corresponding to the capacity limit of the considered channel (depicted in solid line). The SFB in (11) is not applicable for the D_n values above, but only for higher values of D_n. The SFB is shown for $q = 4$ with D_n values of 186 and 1851, respective to the presented block lengths. For $q = 8$, 16 and 32, the two respective curves of the SFB refer to D_n values of 191 and 1951. Except for the quaternary alphabet when large D_n values are used (for which the ϵ_n is around 0.1), for all other cases, the value of ϵ_n ranges from 10^{-12} for a short block length and low constellation order to 10^{-23} for a long block length and higher constellation orders. Notice that the SFB in Theorem 2 deteriorates considerably as compared to the bound provided in Theorem 3, as the constellation order is increased. It is interesting to compare the studied bounds to the union bound which is known to diverged at the cutoff rate of the communication channel. For alphabet cardinalities of $q = 4$ and $q = 8$, the cutoff rate corresponds to $\frac{E_s}{N_0}$ ratios of 2.46 dB and 5.05 dB, respectively, which exemplify the superiority of both derivations over the union bound. However, for alphabet cardinalities of $q = 16$ and $q = 32$, the SFB deteriorate considerably comparing to the bound provided in Theorem 3 and to the union bound which is depicted in plots (c) and (d) (the thresholds corresponding to the cutoff rates for these cases are 7.57 and 10.31 dB, respectively). It is remarkable that the SFB for alphabet size of $q = 32$ performs worse than the union bound. The reason for the deterioration of the SFB is explained when looking into the rate term $\frac{1}{n} \log_q \alpha(\mathcal{C}, D_n)$. This term corresponds to the spectrum difference between the considered ensemble and the fully random-code ensemble, and is depicted in Figure 3 as a function of the ration D_n/n for alphabet sizes of $q = 4$, 8, 16, and 32, and for block lengths of $n = 512$, 1008, and 10080 symbols. The increase of that term, as a function of the alphabet size is cleared, and for $q = 32$ this term rises to more than 10% of the code rate (at D_n ranges where the corresponding ϵ_n value is relevant) which explains the considerable deterioration of the SFB. The bound in Theorem 3 does not present such deterioration because of its derivation which prevents from taking the max-out operation taken in the proof of the SFB in Theorem 2.



(a) $q = 4$



(b) $q = 4$



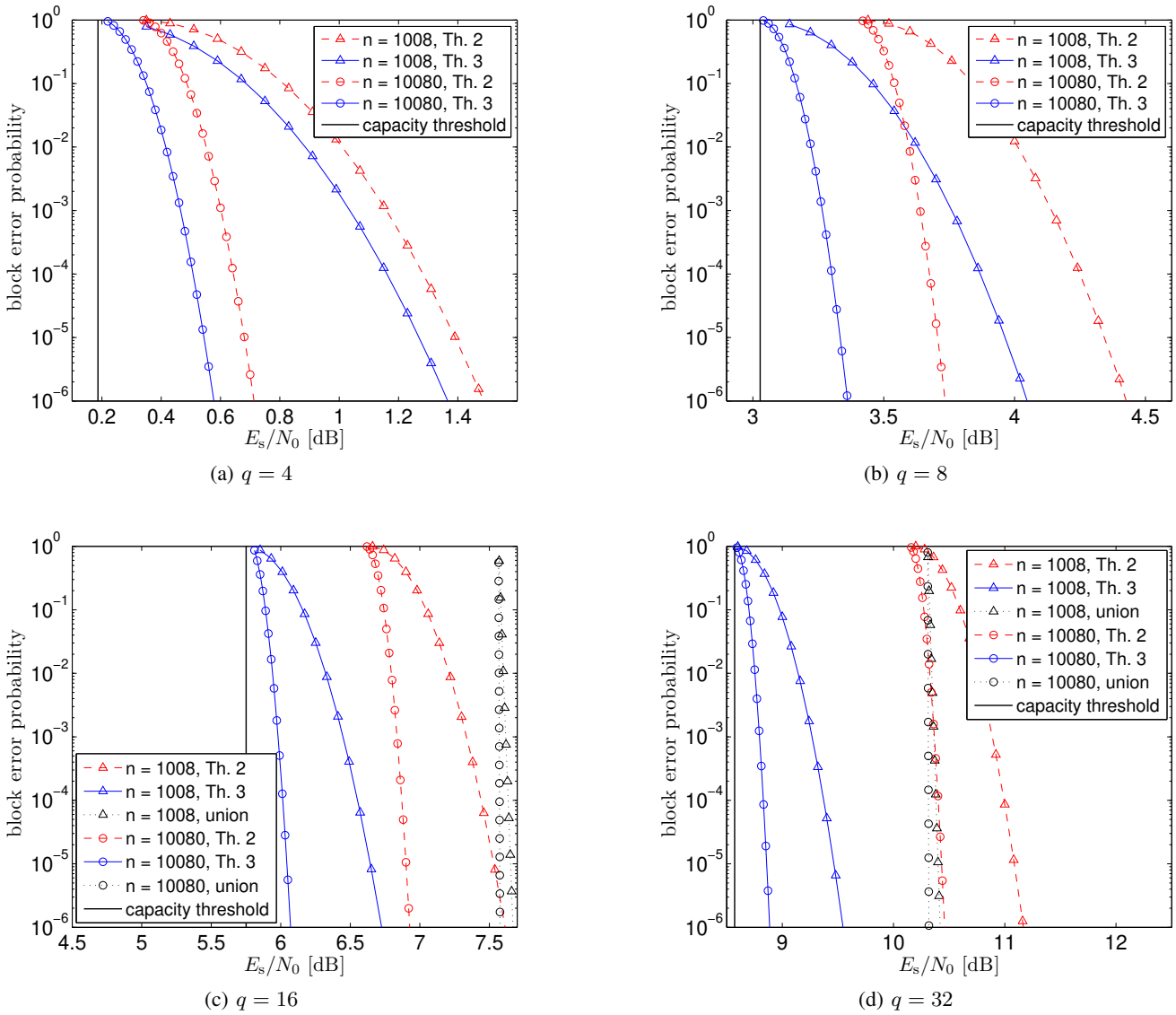
(c) $q = 8$

Fig. 1: Upper bounds on the block error probability under ML decoding of the $(8, 16)$-regular LDPC ensemble of Gallager, whose transmission takes place over the 4-ary and 8-ary symmetric channels; $q = 4$ in plots (a) & (b) and $q = 8$ in plot (c). The upper plot exemplifies the weakness of the union bound, and then plots (b) and (c) focus on the improved bounds. In addition, plots (b) and (c) exemplify a further improvement that is achieved when the bounds are evaluated using the exact spectrum of the code. The capacity limit corresponds to a crossover probability of $p = 0.1893$ and 0.247 for $q = 4$ and 8, respectively (marked as solid vertical lines). This figure depicts the upper bounds on the block error probability for the expurgated ensemble with a block length of 1008 symbols (in square and triangle markers, where the weight spectrum is evaluated using Lemma 3 and Lemma 4, respectively), and $10,080$ symbols (in dot and circle markers, where the weight spectrum is evaluated using Lemma 3 and Lemma 4, respectively), based on Theorems 2 (in dashed lines) and Theorem 3 (in solid lines).

The Divsalar bound [6], [34] is widely used when assessing the error performance of binary turbo-like code ensembles whose transmission takes place over the binary-input AWGN channel (see [1] and references therein). This is due to the fact that the bound is given in a closed form, and its calculation does not involve numerical integration and parameter optimization. The basic concept the bound is based on is the following:

$$\Pr(\text{error}) \quad \leq \quad \Pr(\text{error}, \mathbf{y} \in \mathcal{R}) + \Pr(\mathbf{y} \notin \mathcal{R})$$

where $\mathbf{y}$ is the received vector, and the region $\mathcal{R}$ is the n-dimensional sphere whose center is at a point along the line connecting the origin to the all-zero codeword and whose radius is optimized analytically. In addition, closed form expressions are derived using the Chernoff bound (for the exact details see [1, Chapter 3.2.4]). This technique was generalized by the authors to the non-binary setup by examining various regions in the complex observation



(a) $q = 4$ (b) $q = 8$

(c) $q = 16$ (d) $q = 32$

Fig. 2: Upper bounds on the block error probability under ML decoding of the $(8, 16)$-regular LDPC ensembles of Gallager with alphabet size of $q = 4$, 8, 16, and 32, whose transmission takes place over an AWGN channel with a q-ary PSK modulation. The capacity limit corresponds to an $\frac{E_s}{N_0}$ threshold of 0.19, 3.03, 5.75, and 8.57 dB, respective to the constellation order (marked in vertical solid lines). This figure depicts the upper bounds on the block error probability for the expurgated ensemble with block lengths of 1008 symbols in triangle marker, and $10,080$ symbols in circle marker, based on Theorem 2 (in dashed lines) and Theorem 3 (in solid lines) where the spectrum of the considered ensemble is evaluated using Lemma 4.

space. In a sharp contrast to the binary case, not all the parameters could be optimized analytically. Moreover, the resulting bounds were not satisfactory with respect to the results presented in Example 3 and are therefore omitted.

Example 4 (A Comparison to lower bounds on the decoding error probability)**.** The upper bound in Theorem 3 for the block error probability under ML decoding is compared in Figure 4 to the SP59 lower bound of Shannon [18], and the ISP lower bound in [21]. The regular LDPC code ensembles of Gallager is considered with octal alphabet cardinality and block lengths of 1008 and 10080 symbols, and the performance is studied over the AWGN channel with a 8-PSK modulation. In plot (a) the upper bound provided in Theorem 3 is depicted for the (8,16)-regular expurgated ensemble in solid lines and clear triangle and circle markers, respective to the considered block lengths (the bound is evaluated with the same parameters as in Example 3, Figure 2, plot (b)). In addition, the ultimate performance of a rate 0.5 code is assessed via the SP59 and the ISP lower bounds on the decoding error, depicted with dashed-doted lines with empty markers, and dashed lines with filled marker, respectively (the block length

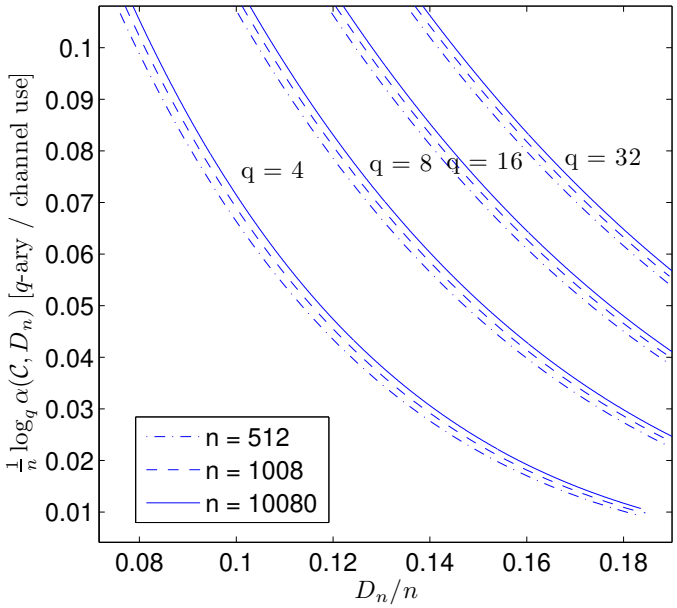
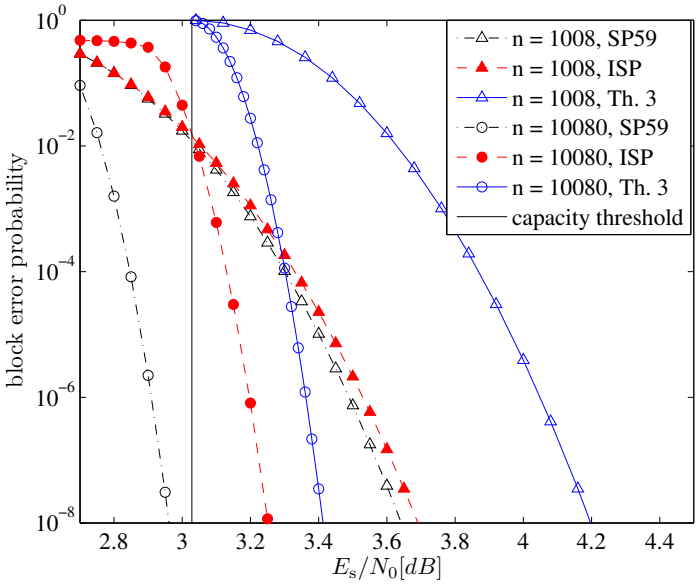


Fig. 3: The term $\frac{1}{n}\log_q \alpha(\mathcal{C}, D_n)$ in (11) for the regular (8,16) LDPC ensemble of Gallager [22], depicted for alphabet sizes of $q = 4$, 8, 16, and 32, and block lengths of $n = 512$, 1008, and 10080 symbols.

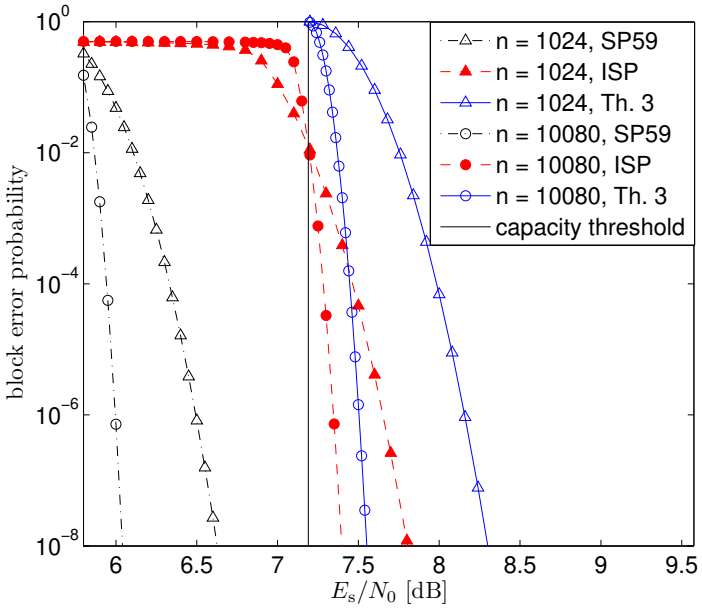
of 1008 is depicted with triangle markers, and the block length of 10080 symbols is depicted with circle markers, as in the case of the compared upper bounds). For a rate 0.5, octal cardinality ensemble and error probability of 10^{-4}, it can be shown that the ISP bound is superior to the SP59 bound above a block length of 863 symbols (and 811 symbols at a block error probability of 10^{-6}). This property of the lower bounds is discussed in details in [21, Section V]. Indeed for the block length of 1008 symbols little difference is observed between the two considered lower bounds, both around 0.5 dB within the respective upper bound in Theorem 3 for all range of interest. For the large block length of 10080 symbols, clear difference is observed between the SP59 and the ISP lower bounds. The ISP is within 0.2 dB from the respective upper bound (in all range of interest error probabilities), and it is more informative than the capacity limit (which corresponds to infinite block length). The SP59 bound on the other hand is less informative than the capacity limit for this block lengths. The difference between the two lower bounds relies on the derivation of the ISP which takes into account the symmetry of the considered modulation whereas the SP59 bound holds only for equal-energy signals. The comparison of the upper bound in Theorem 3 and the considered lower bounds is further studied in plot (b) for (8,32)-regular LDPC code ensembles with block lengths of 1024 and 10080 symbols and whose design rate is 0.75 bits per channel use. The upper bound is depicted in solid lines for an expurgated ensemble with $D_n = 25$ for a block length of 1024 symbols (with triangle markers) and with $D_n = 95$ for a block length of 10080 symbols (with circle markers). For the considered rate and alphabet cardinality, it can be shown that the ISP bound uniformly better than the SP59 bound for all block lengths exceeding 51 symbols at a decoding error probability of 10^{-4} (and 47 symbols for decoding error probability of 10^{-6}) [21]. Indeed it is evident in plot (b), where the ISP bound maintains its close proximity with the provided upper bound. The SP59 bound on the other hand deteriorates considerably for this case, and it is less informative than the capacity limit for both considered block lengths.

IV. Gallager-type bounds for fully-interleaved fading channels with prefect CSI at the receiver

In the section, the error probability of a linear block code $\mathcal{C}$ is considered under ML decoding when transmission takes place over a fully-interleaved fading channel and perfect CSI is available at the receiver. The fading is assumed to be a continuous r.v. (a similar framework is possible for the discrete case). Let $\mathcal{A}$ denote the set of possible fading samples, and $p(\mathbf{y}, \mathbf{a}|\mathbf{x})$ denote the joint pdf of the received sequence $\mathbf{y} = (y_1, \ldots, y_n) \in \mathcal{Y}^n$ and the fading samples $\mathbf{a} = (a_1, \ldots, a_n) \in \mathcal{A}^n$ given that the transmitted codeword is $\mathbf{x} \in \mathcal{C}$. Due to an ideal symbol interleaving,



(a) $R = 0.5$



(b) $R = 0.75$

Fig. 4: A Comparison between the upper bound in Theorem 3 and the SP59 and ISP lower bounds on the decoding error probability for octal alphabet block codes whose transmission takes place over an AWGN channel with 8-ary PSK modulation. This figure depicts the upper and lower bounds on the block error probability for block lengths of 1008 symbols in triangle marker, and $10,080$ symbols in circle marker. The upper bounds (in solid line) are provided for the $(8, 16)$ and $(8, 32)$ regular LDPC expurgated ensembles of Gallager, in plot (a) and (b), respectively. The lower bounds (the SP59 bound in dashed-dotted lines with empty markers and the ISP bound in dashed lines with filled markers), correspond to the ultimate performance of rate–0.5 and rate–0.75 block codes, in plots (a) and (b), respectively. The capacity limits in these cases correspond to $\frac{E_s}{N_0}$ thresholds of 3.03 and 7.19 dB, respective to the ensemble code rate.

the channel is memoryless and accordingly

$$p(\mathbf{y},\mathbf{a}|\mathbf{x}) = \prod_{i=1}^{n} p(y_i|x_i,a_i)p(a_i)$$

where $p(y|x,a)$ is the single-letter conditional pdf of the channel, and $p(a)$ is the pdf of a fading sample. The following definition of symmetry is a generalization to the one presented in Definition 1. This generalization is obtained by directly applying Definition 1 to a channel whose observations are the pair of the considered channel output and the fading sample.

Definition 2. Consider the fully-interleaved fading channel with an input-alphabet $\mathcal{X}$, and perfect CSI at the receiver. The channel, which is characterized by a transition pdf p, is symmetric if for every $a \in \mathcal{A}$, there exists a function $\mathcal{T}_a : \mathcal{Y} \times \mathcal{X} \to \mathcal{Y}$ which satisfies the following properties:

1) For every $x \in \mathcal{X}$ the function $\mathcal{T}_a(\cdot, x) : \mathcal{Y} \to \mathcal{Y}$ is bijective and with a Jacobian 1.
2) For every $x_1, x_2 \in \mathcal{X}$ the following two relations hold

$$p(y|x_1,a) = p(\mathcal{T}_a(y, x_2 - x_1)|x_2, a) \tag{26}$$

and

$$\mathcal{T}_a(\mathcal{T}_a(y,x_1),x_2) = \mathcal{T}_a(y, x_1 + x_2). \tag{27}$$

Notice that this definition of symmetry is a weaker notion compared to a one where there exists a function $\mathcal{T} : \mathcal{Y} \times \mathcal{X} \to \mathcal{Y}$ meeting conditions (26) and (27) for every fading sample $a \in \mathcal{A}$. Nevertheless, this weaker notion is satisfy in order to prove that for the case at hand the ML decoding error probability is not depend on the actual transmitted message. This is clearly expected since Definition 2 is a direct application of Definition 1 for the case at hand. In addition, writing the conditional decoding error probability for the m-th message under ML decoding as

$$P_{\mathrm{e}|m} = \int_{\mathbf{a}} \int_{\mathbf{y} \in \Lambda_m^c(\mathbf{a})} p_{\mathbf{x}_m}(\mathbf{y},\mathbf{a})\, d\mathbf{y}\, d\mathbf{a} = \int_{\mathbf{a}} p(\mathbf{a}) \int_{\mathbf{y} \in \Lambda_m^c(\mathbf{a})} p(\mathbf{y}|\mathbf{x}_m,\mathbf{a})\, d\mathbf{y}\, d\mathbf{a} \tag{28}$$

where $\Lambda_m(\mathbf{a}) \subseteq \mathcal{Y}^n$ is the ML decoding set given that the sequence of fading samples is $\mathbf{a} \in \mathcal{A}^n$, it is enough to show that the inner integral in (28) is independent of the transmitted message m (this is accomplished for every sequence of fading sample sequence $\mathbf{a}$ in the same way as of the proof of Proposition 1).

The following theorem is a generalization of Theorem 1 for the case at hand:

Theorem 4. Consider the transmission of an (n,k) linear block code $\mathcal{C}$ which takes place over a symmetric, fully-interleaved fading channel with perfect CSI at the receiver. Assume that the channel input and output alphabets are $\mathcal{X}$ and $\mathcal{Y}$, respectively, and let p be the transition pdf of the channel. Then, under the assumptions and notation in Lemma 2, the ML decoding error probability satisfies

$$\begin{aligned} P_{\mathrm{e}} \;\leq\; \sum_{j=1}^{J} \Bigg(& \sum_{\mathbf{t} \in \mathcal{T}_j^*:\; n - t_0 > D_n} \mathsf{E}\Big[|\mathcal{C}_{\mathbf{t}}| \;\Big|\; d_{\min} > D_n \Big] \\ & \prod_{x \in \mathcal{X}} \left(\iint_{(y,a)} \psi_j(y,a)^{1-\frac{1}{\rho_j}} p_0(y,a)^{\frac{1-\lambda_j\rho_j}{\rho_j}} p_x(y,a)^{\lambda_j}\, dy\, da \right)^{t_x} \Bigg)^{\rho_j} \;+\; \epsilon_n \end{aligned} \tag{29}$$

where $\left\{\mathcal{T}_j^*\right\}_{j=1}^{J}$ is a division of the entire set of compositions (except for the one of the all zero codeword) to J subsets, $\mathsf{E}\Big[|\mathcal{C}_{\mathbf{t}}| \;\Big|\; d_{\min} > D_n \Big]$ denotes the expectation of the complete composition spectrum under the assumption that $d_{\min} > D_n$, $0 \leq \rho_j \leq 1$ and $\lambda_j \geq 0$ are arbitrary real-valued parameters, and $\psi_j : \mathcal{Y} \times \mathcal{A} \to \mathbb{R}$ are arbitrary non-negative tilting probability measures.

Proof: See Appendix E. ∎

Theorem 4 is applicable with the un-normalized tilting measure used in Theorem 2 and Theorem 3. Moreover, assuming an ensemble satisfying (18) and a division of compositions which depends only on the Hamming weight of the respected compositions, Theorem 4 results in the same bounds as in (11) and (19) where the summations

over $y \in \mathcal{Y}$ are replaced with the appropriate volume integrals over $y \in \mathcal{Y}$ and $a \in \mathcal{A}$. Another, more tedious option is to calculate Theorem 3 in conjunction with the optimum tiling measure. For ensembles satisfying (18), and the particular choice of $J = n$ and $\mathcal{T}_j^* = \{\mathbf{t} :\ n - t_0 = j\}$, using calculus of variations, the optimum tilting measures ψ_j, $Dn < j \leq n$, are given by

$$\psi_j(y,a) = \alpha_{j,0}\, p_0(y,a) \left(1 + \sum_{x \in \mathcal{X}*} \alpha_{j,x} \left(\frac{p_x(y,a)}{p_0(y,a)}\right)^{\lambda_j}\right)^{\rho_j}, \quad \lambda_j \geq 0,\ 0 \leq \rho_j \leq 1$$

where the parameters $\alpha_{j,x}$, $x \in \mathcal{X}^*$ satisfies

$$\alpha_{j,x} = \frac{\frac{j}{n} \iint_{(y,a)} \psi_j(y,a)^{1-\frac{1}{\rho_j}}\, p_0(y,a)^{\frac{1}{\rho_j}}\, dy\, da}{(1-\frac{j}{n}) \sum_{x \in \mathcal{X}^*} \iint_{(y,a)} \psi_j(y,a)^{1-\frac{1}{\rho_j}}\, p_0(y,a)^{\frac{1-\lambda_j \rho_j}{\rho_j}}\, p_x(y,a)^{\lambda_j}\, dy\, da}$$

and $\alpha_{j,0}$ are normalized parameters, which are to be set such that ψ_j are probability measures. The numerical evaluations of such bounds result in a tedious numerical process. It is therefore of interest to seek for tilting measures for which the integration in (29) has a closed form expression. Exponential upper bounds on the ML decoding error probability of binary linear block codes that operate over the binary-input fully-interleaved Rician fading channel with perfect CSI at the receiver were derived in [9]. These bounds are reasonably tight in a certain portion of the rate region exceeding the cutoff rate and do not require numerical integrations involved in the evaluation of the optimal DS2-based bound. In the following example, the technique in [9] is generalized and applied to non-binary linear block codes whose transmission takes place over the fully-interleaved Rician fading channel with a q-ary PSK modulation.

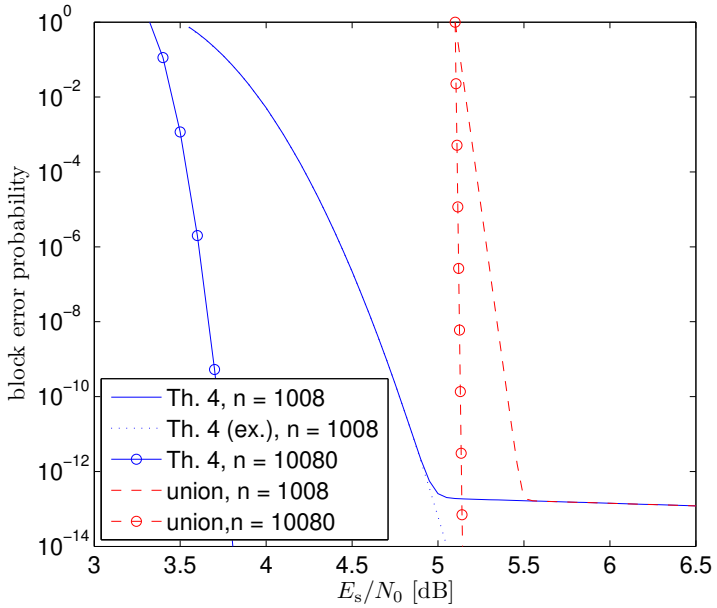
Example 5 (A sub-optimal DS2 bound for the fully-interleaved Rician fading channel with a q-ary PSK modulation)**.** Consider the class of fully-interleaved Rician fading channels with an additive white Gaussian noise. A codeword $\mathbf{x} = (x_1, \ldots, x_n)$ with a block length n and codeword symbols over the alphabet $\mathcal{X} = \{0, 1, \ldots, q-1\}$ is transmitted over a discrete-time, memoryless channel; the received sequence $\mathbf{y} = (y_1, \ldots, y_n) \in \mathbb{C}^n$ satisfies

$$y_k = A_k \sqrt{\frac{2E_{\mathrm{s}}}{N_0}} \exp\left(i \frac{2\pi}{q} x_k\right) + N_k, \quad k = 1, \ldots, n. \tag{30}$$

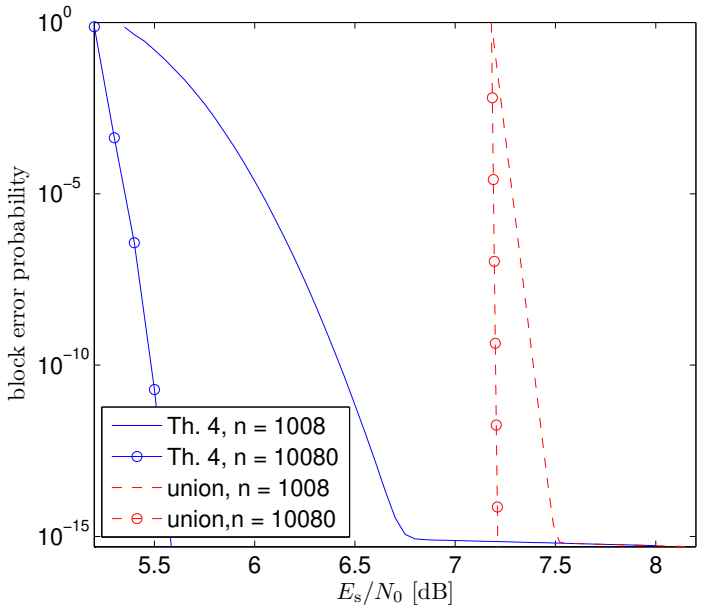
Here A_k is a Rician distributed r.v. with a parameter K, and $N_k = N_k^{\mathrm{r}} + jN_k^{\mathrm{i}}$, where N_k^{r} and N_k^{i} are statistically independent Gaussian r.v.'s with a zero mean and a unit variance. The non-negative real-valued parameter K designates the power ratio between the direct and the diffused paths, $N_0/2$ is the two sided, power density spectrum of the additive white Gaussian noise, and E_{s} is the energy per transmitted coded symbol. The symmetry of the considered channel is guaranteed by the q-ary PSK modulation and the AWGN noise. Following [9], a sub-optimal DS2 bound is suggested for the case at hand. To this end, the exponential tilting measure

$$\psi(y,a) = \frac{\frac{\alpha}{2\pi} \exp\left(-\frac{\alpha}{2}\left|y - au\sqrt{\frac{2E_{\mathrm{s}}}{N_0}}\right|^2 - \frac{\alpha v^2 a^2 E_{\mathrm{s}}}{N_0}\right) p(a)}{\int_0^\infty p(a) \exp\left(-\frac{\alpha v^2 a^2 E_{\mathrm{s}}}{N_0}\right) da}, \quad y \in \mathbb{C},\ a \geq 0 \tag{31}$$

where v and α are non-negative real-valued parameters, and u is a complex-valued parameter. Substituting the exponential tilting measure ψ into (29) provides an upper bound on the error probability which is expressed in a closed form (see Appendix F). The performance of the (8,16) regular non-binary LDPC ensemble of Gallager [22] with block lengths of $n = 1008$ and $n = 10080$ symbols is provided in Figure 5 using the bound in Theorem 4, in solid lines, and the union bound, in dashed lines. The bound in (29) is evaluated with $J = 6$ and the division of compositions is done according to their Hamming weights where barriers are set in 350, 425, 500, 575, and 600 symbols for a block lengths of 1008 symbols (the barriers for the block length of 10080 symbols are set to 3500, 4250, 5000, 5750, and 6000 symbols). The presented performance is for quaternary input-alphabet $q = 4$ and a Rayleigh fading channel, and for octal input-alphabet $q = 8$ and a Rician fading channel with $K = 2$, in plots (a) and (b), respectively. In both plots the non-expurgated ensemble is considered, while in plot (a) the performance for an expurgated ensemble with $D_n = 100$ (with a corresponding $\epsilon_n = 10^{-5}$) is also presented for a block length of 1008 symbols, in a dotted line. In both plots, the union bound diverges bellow the cutoff rate which corresponds



(a) $q = 4, K = 0$



(b) $q = 8, K = 2$

Fig. 5: Upper bounds on the block error probability under ML decoding for the $(8, 16)$-regular LDPC ensemble of Gallager, whose transmission takes place over a fully-interleaved Rician fading channel with a given power ratio K between the direct and the diffused paths, a q-ary PSK modulation, and perfect CSI at the receiver. Plot (a) depicts the performance for $q = 4$ and $K = 0$ (i.e., Rayleigh fading channel), and plot (b) depicts the performance for $q = 8$ and $K = 2$. The performance are evaluated using both the union bound and using the bound provided in Theorem 4 for block lengths of $n = 1008$ and $n = 10080$ symbols. In both plots, the non-expurgated ensemble is considered ($D_n = 0$), while in plot (a) the performance of an expurgated ensemble with $D_n = 100$ is also presented as a comparison for a block length of 1008 symbols.

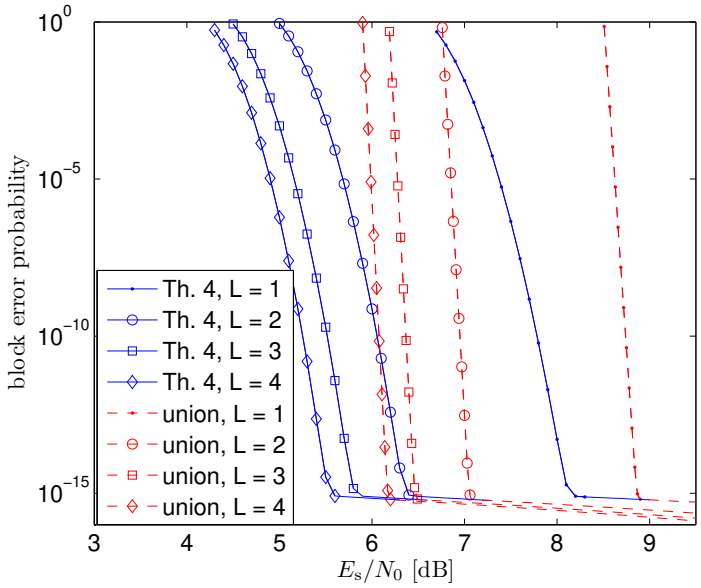


Fig. 6: Upper bounds on the block error probability under ML decoding for the $(8, 16)$-regular LDPC ensemble of Gallager with octal alphabet and a block length of 1008 symbols, whose transmission takes place over the fully-interleaved Rayleigh fading channel with a 8-ary PSK modulation, perfect CSI and MRC diversity combining at the receiver. The figure depicts the performance for MRC diversity combing of $L = 1$ to $L = 4$ antennas at the receiver (with dot, circle, square and diamond markers, respectively). The non-expurgated ensemble is considered ($D_n = 0$), where the performance is evaluated using the bound provided in Theorem 4 (in solid lines) and the union bound (in dashed lines).

to E_s/N_0 thresholds of 5.1 dB and 7.18 dB respectively (the capacity corresponds to thresholds of 1.86 dB and 4.21 dB, respectively). Although the bound in Theorem 4 does not perform (for the considered example) up to the ultimate channel capacity, it is for a block length of 1008 symbols 0.9 dB and 1 dB better than the union bound in plot (a) and 1.2 dB and 1.3 dB in plot (b), at block error rates of 10^{-6}, and 10^{-4}, respectively (for a block length of 10080 symbols the bound in Theorem 4 is better than the union bound by 1.5 dB and 1.8 dB, for quaternary and octal alphabets, respectively, at the considered block error rates).

Example 6 (A sub-optimal DS2 bound for the fully-interleaved Rayleigh fading channel with a q-ary PSK modulation and MRC space diversity)**.** Consider the class of fully-interleaved Rayleigh fading channels with MRC space diversity of order L. The receiver sequence is as in (30) where the fading samples, A_k, are distributed according to the following pdf:

$$p(a) = \frac{2L^L a^{2L-1} \exp\left(-La^2\right)}{(L-1)!}, \ a \geq 0. \tag{32}$$

Note that the E_s/N_0 ration, is normalized here to the point after the MRC module. A closed-form expression for the upper bound on the block error rate, based on Theorem 4 and an exponential tilting measure is suggested (see Appendix 57). The performance of the (8,16) regular non-binary LDPC ensemble of Gallager [22] with octal alphabet and 1008 symbols block length, whose transmission takes place over the considered channel is provided in Figure 6. The bound provided in Theorem 4 is sketched (in solid lines) together with the union bound (in dashed lines), for MRC diversity combining of $L = 1$ to 4 antennas (in dot, circle, square and diamond markers, respectively). Both bounds coincide in the error floor region which is considerably low for the considered ensemble. The union bound is informative only below the cutoff rate, which corresponds to E_s/N_0 of 8.51, 6.76, 6.18, and 5.9 dB, respective to the number of receiving antennas. The bound provided in Theorem 4 is not informative up to the ultimate channel capacity (which corresponds to E_s/N_0 of 4.94, 4.00, 3.68, and 3.30 dB, respective to the number of receiving antennas which is increased from 1 to 4). Nevertheless, the bound in Theorem 4 outperforms the union bound by 1.33 dB and 1.18 dB at block error rates of 10^{-4} and 10^{-6}, respectively, when $L = 1$ receiving antenna is considered, and by 1.02 dB and 1.16 dB, respective to the considered block error rates, when $L = 4$ receiving antennas are used.

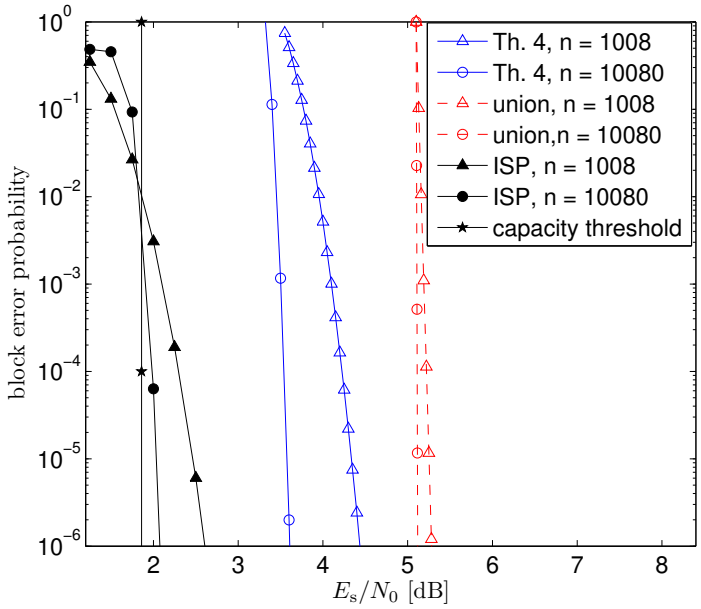


Fig. 7: A comparison between the upper bounds in Example 5 (the sub-optimal DS2 bound with solid lines and the union bound in dashed lines, both with clear markers) on the block error probability under ML decoding for the $(8, 16)$-regular LDPC ensemble of Gallager. The transmission takes place over fully-interleaved Rayleigh fading channel with a QPSK modulation and perfect CSI at the receiver. The ISP lower bounds (with solid line and filled markers) on the decoding error probability are shown for block lengths of 1008 symbols (with triangle markers) and 10080 symbols (with circle markers). The capacity limit for infinite block length is also presented (in a solid line and a star marker as a reference.

Example 7 (A comparison of the upper bounds with the ISP lower bound)**.** The DS2 bound in Theorem 4 is compared in this example to the ISP lower bound provided in [21] on the ultimate error performance of finite-length codes. The bounds are compared for block codes whose transmission takes place over the fully interleaved Rayleigh fading with a quadrature-phase shift-keying (QPSK) modulation and a perfect CSI at the receiver. The DS2 bound is evaluated with the sub-optimal exponential tilting measure in (31), for (8,16)-regular LDPC code ensembles with block lengths of 1008 and 10080 symbols. It is depicted in Figure 7 (the same ensemble and bound parameters as in Example 5 (see Figure 5, plot (a)) are applied; the union bounds for the same code ensembles are also presented in this figure as a reference. The ultimate error performance using a rate–0.5 code with the considered block lengths is evaluated using the ISP lower bound [21]. For the two block lengths considered in this example, the ISP bound is more informative than the capacity threshold for block error rates below 10^{-2}. For a block length of 1008 symbols, the gap between the ISP lower bound and the sub-optimal DS2 upper bound is 2.0 dB and 2.2 dB at block error rates of 10^{-4} and 10^{-6}, respectively. For a block length of 10080 symbols, the gap between the upper and lower bounds is about 1.5 dB for the above block error rates; this considerably closes the gap between the union upper bounds and the respective ISP lower bounds (which is equal to 3.0 dB).

V. Summary and Conclusions

This paper considers Gallager-type upper bounds on the decoding error probability of non-binary linear block codes whose transmission takes place over memoryless symmetric channels. The general bounding approach is based on a partitioning of the original ensemble into two subsets of codebooks according to their minimal Hamming distance. It is combined of two bounds where the performance of the set of codebooks with a relatively low minimum Hamming distance is assessed via a simple union bound, whereas the bound for the other set of codebooks is evaluated using the second version of the Duman and Salehi (DS2) bound (for a presentation of the DS2 bound, the reader is referred to [1]). As a particular case of this bounding technique, an adaptation of the Shulman-Feder bound (SFB) (see [10]) is provided for non-binary linear block codes. The latter approach which is related to the adaptation of the SFB for the non-binary setting is similar to the one provided by Bennatan and Burshtein [11] for a different setting of coding with a random coset mechanism. Under a symmetry property of the ensemble, the

resulting bound is considerably simplified and even tightened. This simplifying assumption, which holds in particular for the considered non-binary low-density parity-check (LDPC) ensembles, yields a bound whose summations are over the Hamming weights of the non-zero codewords rather than their compositions (see Theorem 3).

The tightness of the bounds presented in this paper is exemplified for the non-binary regular LDPC ensembles of Gallager [22]. The exact complete composition spectrum is provided for these ensembles by generalizing the derivation in [32] which applies to the binary setting. Performance analysis is studied assuming that the transmission takes place over memoryless symmetric channels, e.g., the q-ary symmetric channel, the additive white Gaussian noise (AWGN) channel with a q-ary phase-shift keying (PSK) modulation, and fully interleaved fading channels with perfect channel state information (CSI) at the receiver. The tightening of the bounds in Theorems 2 and 3 for expurgated non-binary regular LDPC code ensembles, due to the replacement of the upper bound on the complete composition spectrum of these ensembles (as provided in [2]) with its exact expression, is exemplified in this paper.

The bound provided in Theorem 3 is most attractive and shows meaningful results at a significant portion of the rate region between the cutoff rate and the ultimate channel capacity. The weakness of the union bound, which diverges at the cutoff rate for long enough block codes, is the basic motivation for the replacement of the union bound with some improved upper bounds on the decoding error probability (as done in this paper). The adaptation of the SFB in Theorem 2 is not as informative as the one in Theorem 3, and deteriorates considerably as the alphabet size increases. This deterioration relates to the derivation of the bound which on one hand reproduces the 1965 Gallager bound for random coding, but on the other hand has an additional term in the error exponent which corresponds to the deviation of the average distance spectrum of the code ensemble from the binomial distribution (where the latter distribution characterizes the average distance spectrum of the ensemble of fully random block codes). Unlike the asymptotic analysis (both in terms of the block lengths and the node degrees) provided in [11], for the non-binary regular LDPC ensemble of Gallager, as considered in this paper, this additional term in the error exponent of the SFB has a considerable effect on the SFB variation studied.

The upper bound in Theorem 3 is compared to two lower bounds on the ultimate error performance of finite-length block codes (which hold for general block codes, either linear or non-linear): The 1959 sphere-packing (SP59) lower bound of Shannon [18], and the improved sphere-packing (ISP) lower bound derived in [21][2]. The comparison between the upper bounds derived in this paper and the above sphere-packing lower bounds is exemplified for various memoryless and symmetric channels. These comparisons show that the ISP bound is a useful lower bound on the decoding error probability of finite-length block codes (or ensembles), and this bound is often more informative than the capacity limit and shows in some cases close proximity to the compared upper bounds (see also [21, Section 5]).

The Divsalar bound [6], [34] is widely used when assessing the error performance of binary turbo-like code ensembles whose transmission takes place over the binary-input AWGN channel (see [1, Chapter 3.2.4] and references therein). This technique was generalized by the authors to the non-binary setup by examining various geometric interpretations of this bound. In contrast to the binary case, not all the parameters could be optimized analytically. Moreover, the resulting bounds were rather loose as compared to the bounds presented in this paper, and we therefore omit the derivation of these bounds in this paper.

A sub-optimal variation of the DS2 bound is derived for fully interleaved fading channels with perfect CSI at the receiver. This bound generalizes the results in [9] (in terms of the alphabet cardinality). In particular, a sub-optimal exponential tilting measure is suggested which results in closed form expressions (which are subjected to a numerical optimization). The resulting bounds are not informative up to the ultimate capacity limit. However, they considerably outperform the union bound. The sub-optimal upper bound is compared to the ISP lower bound, and this comparison further exemplifies the improvement of the provided upper bound over the union bound. An alternative for the upper bound considered in this paper for fully interleaved fading channels with perfect CSI at the receiver is to use Theorem 4 in conjunction with optimal tilting measures. This may be accomplished in a similar way to the one derived in [8] for memoryless binary-input output-symmetric (MBIOS) channels. The mathematical structure of such tilting measures was derived by the authors, though the required numerical evaluation seems to be too tedious and time consuming.

A derivation of the bounds presented in this paper for the case of transmission of non-binary codes which

[2]This term is coined in [21] and relates to the improvement of the 1967 sphere-packing bound of Shannon, Gallager and Berlakamp [19] and the sphere-packing lower bound in [20] for finite block length whose transmission takes place over memoryless *symmetric* channels.

takes place over parallel channels forms a natural continuation to this work. This may be accomplished using the bounding techniques provided in [16] and [17] for the case of MBIOS channels. Such a generalization is in order when considering the ultimate performance of modern coded multi-carrier systems.

All bounds provided in this paper are easily applicable for the regular LDPC ensembles provided in [11] (although asymptotic performance was studied in [11], upper bounds on the complete composition spectrum of the considered ensembles are available in [11] for finite block lengths). It is very appealing to evaluate the decoding error performance of irregular non-binary LDPC ensembles and other non-binary turbo-like ensembles (such as non-binary repeat-accumulate ensembles) where this requires further study of the complete composition spectra of these code ensembles.

Acknowledgment

The authors are grateful to Dr. Gil Wiechman for providing his Matlab computer programs for the calculation of the ISP bounds (this software was written as part of his joint work with the second co-author on improved sphere-packing bounds [21]).

APPENDIX A
PROOF OF PROPOSITION 1

The following proof holds for channels with a discrete-output alphabet, and the generalization of the proof to continuous-output alphabet channels is trivial. Let p be the symmetric transition probability function of the considered channel, and $\mathcal{T}$ be its corresponding function according to Definition 1. The conditional error probability of the m-th message, $\mathbf{x}_m = (x_{m,1}, x_{m,2}, \ldots, x_{m,n})$, under ML decoding is given by

$$\begin{aligned} P_{\mathrm{e}|m} &= \sum_{\mathbf{y}\in\Lambda_m^c} \prod_{i=1}^{n} p\left(y_i|x_{m,i}\right) = \sum_{\mathbf{y}\in\Lambda_m^c} \prod_{x\in\mathcal{X}} \prod_{\{i:\ x_{m,i}=x\}} p(y_i|x) \\ &= \sum_{\mathbf{y}\in\Lambda_m^c} \prod_{x\in\mathcal{X}} \prod_{\{i:\ x_{m,i}=x\}} p(\mathcal{T}\left(y_i,-x\right)|0) \end{aligned}$$

where $\mathbf{y} = (y_1, \ldots, y_n)$, and

$$\begin{aligned} \Lambda_m^c &= \left\{ \mathbf{y} : \sum_{i=1}^{n} \ln\left(\frac{p(y_i|x_{m',i})}{p(y_i|x_{m,i})}\right) \geq 0,\ \text{for some } m' \neq m \right\} \qquad (33) \\ &= \left\{ \mathbf{y} : \sum_{\{x,x'\in\mathcal{X}:\ x'\neq x\}} \sum_{\{i:\ x_{m',i}=x',x_{m,i}=x\}} \ln\left(\frac{p(y_i|x')}{p(y_i|x)}\right) \geq 0,\ \text{for some } m' \neq m \right\} \\ &= \left\{ \mathbf{y} : \sum_{\{x,x'\in\mathcal{X}:\ x'\neq x\}} \sum_{\{i:\ x_{m',i}=x',x_{m,i}=x\}} \ln\left(\frac{p(\mathcal{T}(y_i,-x')|0)}{p(\mathcal{T}(y_i,-x)|0)}\right) \geq 0,\ \text{for some } m' \neq m \right\}. \end{aligned}$$

Using the change of variables

$$z_i = \mathcal{T}(y_i, -x_{m,i}), \quad 1 \leq i \leq n$$

it follows that

$$P_{\mathrm{e}|m} = \sum_{\mathbf{z}\in\tilde{\Lambda}_m^c} \prod_{i=1}^{n} p(z_i|0)$$

where

$$\begin{aligned} \tilde{\Lambda}_m^c &= \left\{ \mathbf{z} : \sum_{\{x,x'\in\mathcal{X}:\ x'\neq x\}} \sum_{\{i:\ x_{m',i}=x',x_{m,i}=x\}} \ln\left(\frac{p(\mathcal{T}(z_i,x-x')|0)}{p(z_i|0)}\right) \geq 0,\ \text{for some } m' \neq m \right\} \\ &= \left\{ \mathbf{z} : \sum_{\delta\in\mathcal{X}} \sum_{\{i:\ x_{m,i}-x_{m',i}=\delta\}} \ln\left(\frac{p(\mathcal{T}(z_i,\delta)|0)}{p(z_i|0)}\right) \geq 0,\ \text{for some } m' \neq m \right\}. \end{aligned}$$

Since the code $\mathcal{C}$ is a closed linear space, then for every two codewords $\mathbf{x}_{m'} \neq \mathbf{x}_m$ in $\mathcal{C}$, there exists a third non-zero codeword $\mathbf{x}_l$ in $\mathcal{C}$ where $\mathbf{x}_l = \mathbf{x}_{m'} - \mathbf{x}_m$. Hence, for every $m = 1, 2, \ldots, M$ and every $\mathbf{z} \in \tilde{\Lambda}_m^c$, there exists some $l \in \{1, 2, \ldots, M\}$ for which

$$\sum_{\delta \in \mathcal{X}} \sum_{\{i:\ -x_{l,i} = \delta\}} \ln\left(\frac{p(\mathcal{T}(z_i, \delta)|0)}{p(z_i|0)}\right) \geq 0.$$

Denote by $\mathbf{x}_1 \in \mathcal{C}$ the all-zero codeword, then it follows that

$$\tilde{\Lambda}_m^c = \tilde{\Lambda}_1^c, \quad m = 1, 2, \ldots, q^k \tag{34}$$

which concludes the proof.

APPENDIX B
PROOF OF PROPOSITION 2

Since the channel is symmetric, we have from Proposition 1 and (3) that

$$\begin{aligned} P_{\mathrm{e}} \;=\; P_{\mathrm{e}|0} \leq & \left(\sum_{\mathbf{y} \in \mathcal{Y}^n} G_n^0(\mathbf{y}) p_n(\mathbf{y}|\mathbf{0})\right)^{1-\rho} \\ & \cdot \left\{\sum_{m' \neq 0} \sum_{\mathbf{y} \in \mathcal{Y}^n} G_n^0(\mathbf{y})^{1-\frac{1}{\rho}} p_n(\mathbf{y}|\mathbf{0}) \left(\frac{p_n(\mathbf{y}|\mathbf{x}_{m'})}{p_n(\mathbf{y}|\mathbf{0})}\right)^{\lambda}\right\}^{\rho}. \end{aligned} \tag{35}$$

Next, setting $G_n^0(\mathbf{y})$ as in (4), for memoryless channels we have

$$P_{\mathrm{e}} \;\leq\; \left(\sum_{\mathbf{y} \in \mathcal{Y}^n} \prod_{i=1}^{n} g(y_i) p(y_i|0)\right)^{1-\rho} \cdot \left\{\sum_{m' \neq 0} \sum_{\mathbf{y} \in \mathcal{Y}^n} \prod_{i=1}^{n} g(y_i)^{1-\frac{1}{\rho}} p(y_i|0) \left(\frac{p(y_i|x_{m',i})}{p(y_i|0)}\right)^{\lambda}\right\}^{\rho} \tag{36}$$

which concludes the proof by replacing the sum of products with the correspondent product of sums.

APPENDIX C
COMPLEMENTARY DETAILS REGARDING THE PROOF OF THEOREM 2

The following two technical Lemmas are applied in the derivation of (15):

Lemma 5. Setting (14) in the summation $\sum_{y \in \mathcal{Y}} g(y)^{\xi} p(y|0)^t$ results in

$$\sum_{y \in \mathcal{Y}} \left[\left(\frac{1}{q} \sum_{x \in \mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{\xi\rho} \cdot \left(\frac{1}{q} \sum_{x \in \mathcal{X}} p(y|x)^{t - \frac{\xi\rho}{1+\rho}}\right)\right] \tag{37}$$

for all ξ and t.

Proof: Since p is symmetric, then there exists a function $\mathcal{T}$, as in Definition 1, satisfying (1) and (2). As a

result, setting $g(y)$ as in (14) we have

$$\begin{aligned}
&\sum_{y\in\mathcal{Y}} g(y)^{\xi} p(y|0)^t \\
&= \sum_{y\in\mathcal{Y}} \left(\left(\frac{1}{q} \sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}} \right)^{\rho} p(y|0)^{-\frac{\rho}{1+\rho}} \right)^{\xi} p(y|0)^t \\
&= \sum_{y\in\mathcal{Y}} p(y|0)^{t-\frac{\xi\rho}{1+\rho}} \left(\frac{1}{q} \sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}} \right)^{\xi\rho} \\
&\overset{\text{(a)}}{=} \frac{1}{q} \sum_{x'\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(y|0)^{t-\frac{\xi\rho}{1+\rho}} \left(\frac{1}{q} \sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}} \right)^{\xi\rho} \\
&\overset{\text{(b)}}{=} \frac{1}{q} \sum_{x'\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(\mathcal{T}(y,x')|x')^{t-\frac{\xi\rho}{1+\rho}} \left(\frac{1}{q} \sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}} \right)^{\xi\rho} \\
&\overset{\text{(c)}}{=} \frac{1}{q} \sum_{x'\in\mathcal{X}} \sum_{y'\in\mathcal{Y}} p(y'|x')^{t-\frac{\xi\rho}{1+\rho}} \left(\frac{1}{q} \sum_{x\in\mathcal{X}} p(\mathcal{T}(y',-x')|x)^{\frac{1}{1+\rho}} \right)^{\xi\rho}
\end{aligned}$$

where in (a) an additional variable is added, (b) is based on (1), and (c) follows since

$$\mathcal{T}(\mathcal{T}(y,x)) = y \tag{38}$$

for every $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. Notice how the change of variables in (c) requires the bijectivity of $\mathcal{T}(\cdot, x)$ for every $x \in \mathcal{X}$. Next, using the closure of the (finite) input alphabet, it follows that

$$\begin{aligned}
&\sum_{y\in\mathcal{Y}} g(y)^{\xi} p(y|0)^t \\
&\overset{\text{(a)}}{=} \frac{1}{q} \sum_{x'\in\mathcal{X}} \sum_{y'\in\mathcal{Y}} p(y'|x')^{t-\frac{\xi\rho}{1+\rho}} \left(\frac{1}{q} \sum_{x\in\mathcal{X}} p(\mathcal{T}(\mathcal{T}(y',-x'),x+x'-x)|x+x')^{\frac{1}{1+\rho}} \right)^{\xi\rho} \\
&\overset{\text{(b)}}{=} \frac{1}{q} \sum_{x'\in\mathcal{X}} \sum_{y'\in\mathcal{Y}} p(y'|x')^{t-\frac{\xi\rho}{1+\rho}} \left(\frac{1}{q} \sum_{x\in\mathcal{X}} p(y'|x+x')^{\frac{1}{1+\rho}} \right)^{\xi\rho} \\
&= \frac{1}{q} \sum_{x'\in\mathcal{X}} \sum_{y'\in\mathcal{Y}} p(y'|x')^{t-\frac{\xi\rho}{1+\rho}} \left(\frac{1}{q} \sum_{x''\in\mathcal{X}} p(y'|x'')^{\frac{1}{1+\rho}} \right)^{\xi\rho} \\
&= \sum_{y'\in\mathcal{Y}} \left[\left(\frac{1}{q} \sum_{x'\in\mathcal{X}} p(y'|x')^{t-\frac{\xi\rho}{1+\rho}} \right) \cdot \left(\frac{1}{q} \sum_{x''\in\mathcal{X}} p(y'|x'')^{\frac{1}{1+\rho}} \right)^{\xi\rho} \right]
\end{aligned}$$

where (a) follows from (1) and (b) follows from (2) and (38), both with $x_1 = x$ and $x_2 = x + x'$. This concludes the proof. ■

Lemma 6. For every $\tilde{x} \in \mathcal{X}$,

$$\begin{aligned}
&\sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}^*} p(y|x)^{\frac{1}{1+\rho}} \left(\frac{1}{q} \sum_{x'\in\mathcal{X}} p(y|x')^{\frac{1}{1+\rho}} \right)^{\rho-1} p(y|0)^{\frac{1}{1+\rho}} \\
&\qquad = \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}\setminus\{\tilde{x}\}} p(y|x)^{\frac{1}{1+\rho}} \left(\frac{1}{q} \sum_{x'\in\mathcal{X}} p(y|x')^{\frac{1}{1+\rho}} \right)^{\rho-1} p(y|\tilde{x})^{\frac{1}{1+\rho}}
\end{aligned} \tag{39}$$

where $\mathcal{X}^*$ designates the input alphabet $\mathcal{X}$ with the additive identity element, $0 \in \mathcal{X}$, excluded.

Proof: Since p is symmetric, then there exists a function $\mathcal{T}$, as in Definition 1, satisfying (1) and (2). As a result, (1) implies that for every $\tilde{x} \in \mathcal{X}$

$$\sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}^*} p(y|x)^{\frac{1}{1+\rho}} \left(\frac{1}{q}\sum_{x'\in\mathcal{X}} p(y|x')^{\frac{1}{1+\rho}}\right)^{\rho-1} p(y|0)^{\frac{1}{1+\rho}}$$
$$= \sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}^*} p(y|x)^{\frac{1}{1+rho}} \left(\frac{1}{q}\sum_{x'\in\mathcal{X}} p(y|x')^{\frac{1}{1+\rho}}\right)^{\rho-1} p(\mathcal{T}(y,\tilde{x})|\tilde{x})^{\frac{1}{1+\rho}}. \quad (40)$$

Next, setting $y' = \mathcal{T}(y,\tilde{x})$, from (38) it follows that $y = \mathcal{T}(y',-\tilde{x})$, and the equality in (40) can be rewritten as

$$\sum_{y'\in\mathcal{Y}} \left(\sum_{x\in\mathcal{X}^*} p(\mathcal{T}(y',-\tilde{x})|x)^{\frac{1}{1+\rho}}\right) \left(\frac{1}{q}\sum_{x'\in\mathcal{X}} p(\mathcal{T}(y',-\tilde{x})|x')^{\frac{1}{1+\rho}}\right)^{\rho-1} p(y'|\tilde{x})^{\frac{1}{1+\rho}}$$
$$\stackrel{\text{(a)}}{=} \sum_{y'\in\mathcal{Y}} \left(\sum_{x\in\mathcal{X}^*} p(\mathcal{T}(\mathcal{T}(y',-\tilde{x}),x+\tilde{x}-x)|x+\tilde{x})^{\frac{1}{1+\rho}}\right)$$
$$\left(\frac{1}{q}\sum_{x'\in\mathcal{X}} p(\mathcal{T}(\mathcal{T}(y',-\tilde{x}),x'+\tilde{x}-x')|x'+\tilde{x})^{\frac{1}{1+\rho}}\right)^{\rho-1} p(y'|\tilde{x})^{\frac{1}{1+\rho}}$$
$$\stackrel{\text{(b)}}{=} \sum_{y'\in\mathcal{Y}} \left(\sum_{x\in\mathcal{X}^*} p(\mathcal{T}(y',-\tilde{x}+x+\tilde{x}-x)|x+\tilde{x})^{\frac{1}{1+\rho}}\right)$$
$$\left(\frac{1}{q}\sum_{x'\in\mathcal{X}} p(\mathcal{T}(y',-\tilde{x}+x'+\tilde{x}-x')|x'+\tilde{x})^{\frac{1}{1+\rho}}\right)^{\rho-1} p(y'|\tilde{x})^{\frac{1}{1+\rho}}$$
$$= \sum_{y'\in\mathcal{Y}} \left(\sum_{x\in\mathcal{X}^*} p(y'|x+\tilde{x})^{\frac{1}{1+\rho}}\right) \left(\frac{1}{q}\sum_{x'\in\mathcal{X}} p(y'|x'+\tilde{x})^{\frac{1}{1+\rho}}\right)^{\rho-1} p(y'|\tilde{x})^{\frac{1}{1+\rho}}$$
$$= \sum_{y'\in\mathcal{Y}} \left(\sum_{\tilde{x}'\in\mathcal{X}\setminus\{\tilde{x}\}} p(y'|\tilde{x}')^{\frac{1}{1+\rho}}\right) \left(\frac{1}{q}\sum_{\tilde{x}''\in\mathcal{X}} p(y'|\tilde{x}'')^{\frac{1}{1+\rho}}\right)^{\rho-1} p(y'|\tilde{x})^{\frac{1}{1+\rho}}$$

where (a) follows from (1) with $x_1 = x$ and $x_2 = x + \tilde{x}$ for the first term and with, $x_1 = x'$ and $x_2 = x' + \tilde{x}$ for the second term, and (b) follows from (2) with $x_1 = -\tilde{x}$ and $x_2 = x + \tilde{x} - x$ for the first term and with, $x_1 = -\tilde{x}$ and $x_2 = x' + \tilde{x} - x'$ for the second term. This concludes the proof. Again, notice how the change of variables requires the bijectivity of $\mathcal{T}(\cdot, x)$ for every $x \in \mathcal{X}$. ■

Following is a complete derivation of the setting of $g(y)$ as specified in (14). From (14) and (37), it follows that

$$\begin{aligned}
&\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}} p(y|0)\left(\frac{p(y|x)}{p(y|0)}\right)^{\lambda} \\
&= \sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}} p(y|0) + \sum_{x\in\mathcal{X}^*}\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}} p(y|0)\left(\frac{p(y|x)}{p(y|0)}\right)^{\lambda} \\
&\overset{\text{(a)}}{=} \sum_{y\in\mathcal{Y}}\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{\left(1-\frac{1}{\rho}\right)\rho}\cdot\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{1-\frac{\left(1-\frac{1}{\rho}\right)\rho}{1+\rho}}\right) \\
&\qquad + \sum_{x\in\mathcal{X}^*}\sum_{y\in\mathcal{Y}}\left(\left(\frac{1}{q}\sum_{x'\in\mathcal{X}} p(y|x')^{\frac{1}{1+\rho}}\right)^{\rho} p(y|0)^{-\frac{\rho}{1+\rho}}\right)^{1-\frac{1}{\rho}} p(y|0)\left(\frac{p(y|x)}{p(y|0)}\right)^{\lambda} \\
&= \sum_{y\in\mathcal{Y}}\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{\rho-1}\cdot\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{2}{1+\rho}}\right) \\
&\qquad + \sum_{x\in\mathcal{X}^*}\sum_{y\in\mathcal{Y}}\left(\frac{1}{q}\sum_{x'\in\mathcal{X}} p(y|x')^{\frac{1}{1+\rho}}\right)^{\rho-1} p(y|0)^{1-\frac{\rho}{1+\rho}\left(1-\frac{1}{\rho}\right)-\frac{1}{1+\rho}} p(y|x)^{\frac{1}{1+\rho}} \\
&= \sum_{y\in\mathcal{Y}}\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{\rho-1}\cdot\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{2}{1+\rho}}\right) \\
&\qquad + \sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}^*}\left(\frac{1}{q}\sum_{x'\in\mathcal{X}} p(y|x')^{\frac{1}{1+\rho}}\right)^{\rho-1} p(y|0)^{\frac{1}{1+\rho}} p(y|x)^{\frac{1}{1+\rho}}. \qquad (41)
\end{aligned}$$

where the first term in (a) follows from (37) with $\xi = 1-\frac{1}{\rho}$ and $t=1$, the second term in (a) follows by setting $g(y)$ as in (14). Using (39), we have

$$\begin{aligned}
&\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}} p(y|0)\left(\frac{p(y|x)}{p(y|0)}\right)^{\lambda} \\
&= \sum_{y\in\mathcal{Y}}\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{\rho-1}\cdot\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{2}{1+\rho}}\right) \\
&\quad + \frac{1}{q}\sum_{\tilde{x}\in\mathcal{X}}\sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}\setminus\{\tilde{x}\}}\left(\frac{1}{q}\sum_{x'\in\mathcal{X}} p(y|x')^{\frac{1}{1+\rho}}\right)^{\rho-1} p(y|\tilde{x})^{\frac{1}{1+\rho}} p(y|x)^{\frac{1}{1+\rho}} \qquad (42) \\
&= \sum_{y\in\mathcal{Y}}\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{\rho-1}\frac{1}{q}\left(\sum_{x\in\mathcal{X}} p(y|x)^{\frac{2}{1+\rho}} + \sum_{\tilde{x}\in\mathcal{X}}\sum_{x\in\mathcal{X}\setminus\{\tilde{x}\}} p(y|\tilde{x})^{\frac{1}{1+\rho}} p(y|x)^{\frac{1}{1+\rho}}\right) \\
&= \sum_{y\in\mathcal{Y}}\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{\rho-1} q\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{2} \\
&= q\sum_{y\in\mathcal{Y}}\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}. \qquad (43)
\end{aligned}$$

In addition,

$$\begin{aligned}
&\sum_{x\in\mathcal{X}\setminus\{0\}}\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}}p(y|0)\left(\frac{p(y|x)}{p(y|0)}\right)^{\lambda}\\
&=\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}}p(y|0)\left(\frac{p(y|x)}{p(y|0)}\right)^{\lambda}-\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}}p(y|0)\\
&\overset{\text{(a)}}{=}\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}}p(y|0)\left(\frac{p(y|x)}{p(y|0)}\right)^{\lambda}-\sum_{y\in\mathcal{Y}}\left(\frac{1}{q}\sum_{x\in\mathcal{X}}p(y|x)^{\frac{1}{1+\rho}}\right)^{\rho-1}\left(\frac{1}{q}\sum_{x\in\mathcal{X}}p(y|x)^{\frac{2}{1+\rho}}\right)\\
&\overset{\text{(b)}}{=}q\sum_{y\in\mathcal{Y}}\left(\frac{1}{q}\sum_{x\in\mathcal{X}}p(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}-\sum_{y\in\mathcal{Y}}\left(\frac{1}{q}\sum_{x\in\mathcal{X}}p(y|x)^{\frac{1}{1+\rho}}\right)^{\rho-1}\left(\frac{1}{q}\sum_{x\in\mathcal{X}}p(y|x)^{\frac{2}{1+\rho}}\right) \qquad (44)
\end{aligned}$$

where (a) follows from Lemma 5 with $\xi = 1-\frac{1}{\rho}$ and $t=1$, and (b) follows from (43). Next, from (43) and (44), it follows that

$$\begin{aligned}
&\left[\sum_{l=D_n+1}^{n}\binom{n}{l}\left(\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}}p(y|0)\right)^{n-l}\left(\sum_{x\in\mathcal{X}^*}\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}}p(y|0)^{1-\lambda}p(y|x)^{\lambda}\right)^{l}\right]^{\rho}\\
&\quad\cdot\left(\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}}p(y|0)^{1-\lambda}p(y|x)^{\lambda}\right)^{-n\rho}\\
&=\left[\sum_{l=D_n+1}^{n}\binom{n}{l}\left(\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}}p(y|0)\right)^{n-l}\left(q\sum_{y\in\mathcal{Y}}\left(\frac{1}{q}\sum_{x\in\mathcal{X}}p(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}\right.\right.\\
&\qquad\left.\left.-\sum_{y\in\mathcal{Y}}\left(\frac{1}{q}\sum_{x\in\mathcal{X}}p(y|x)^{\frac{1}{1+\rho}}\right)^{\rho-1}\left(\frac{1}{q}\sum_{x\in\mathcal{X}}p(y|x)^{\frac{2}{1+\rho}}\right)\right)^{l}\right]^{\rho}\\
&\quad\cdot\left(q\sum_{y\in\mathcal{Y}}\left(\frac{1}{q}\sum_{x\in\mathcal{X}}p(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}\right)^{-n\rho}\\
&\overset{\text{(a)}}{=}\left[\sum_{l=D_n+1}^{n}\binom{n}{l}\left(\sum_{y\in\mathcal{Y}}\left(\frac{1}{q}\sum_{x\in\mathcal{X}}p(y|x)^{\frac{1}{1+\rho}}\right)^{\rho-1}\left(\frac{1}{q}\sum_{x\in\mathcal{X}}p(y|x)^{\frac{2}{1+\rho}}\right)\right)^{n-l}\right.\\
&\qquad\left.\left(q\sum_{y\in\mathcal{Y}}\left(\frac{1}{q}\sum_{x\in\mathcal{X}}p(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}-\sum_{y\in\mathcal{Y}}\left(\frac{1}{q}\sum_{x\in\mathcal{X}}p(y|x)^{\frac{1}{1+\rho}}\right)^{\rho-1}\left(\frac{1}{q}\sum_{x\in\mathcal{X}}p(y|x)^{\frac{2}{1+\rho}}\right)\right)^{l}\right]^{\rho}\\
&\quad\cdot\left(q\sum_{y\in\mathcal{Y}}\left(\frac{1}{q}\sum_{x\in\mathcal{X}}p(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}\right)^{-n\rho} \qquad (45)
\end{aligned}$$

where (a) follows from Lemma 5 with $\xi = 1 - \frac{1}{\rho}$ and $t = 1$. Define

$$\begin{aligned}\beta(\rho, D_n) \triangleq & \left(\sum_{y\in\mathcal{Y}}\left(\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}\right)^{-n\rho} \\ & \cdot \left[\sum_{l=D_n+1}^{n}\binom{n}{l}\left(\sum_{y\in\mathcal{Y}}\left(\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{\rho-1}\left(\sum_{x\in\mathcal{X}} p(y|x)^{\frac{2}{1+\rho}}\right)\right)^{n-l}\right. \\ & \left.\left(\sum_{y\in\mathcal{Y}}\left(\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho} - \sum_{y\in\mathcal{Y}}\left(\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{\rho-1}\left(\sum_{x\in\mathcal{X}} p(y|x)^{\frac{2}{1+\rho}}\right)\right)^{l}\right]^{\rho}. \end{aligned} \tag{46}$$

we observe that $\beta(\rho, D_n) < 1$ for every $0 \le D_n \le n$. This is easily verified since

$$\begin{aligned} & \left[\sum_{l=D_n+1}^{n}\binom{n}{l}\left(\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}} p(y|0)\right)^{n-l}\left(\sum_{x\in\mathcal{X}^*}\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}} p(y|0)^{1-\lambda} p(y|x)^{\lambda}\right)^{l}\right]^{\rho} \\ & < \left[\sum_{l=0}^{n}\binom{n}{l}\left(\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}} p(y|0)\right)^{n-l}\left(\sum_{x\in\mathcal{X}^*}\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}} p(y|0)^{1-\lambda} p(y|x)^{\lambda}\right)^{l}\right]^{\rho} \\ & = \left(\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}} p(y|0)^{1-\lambda} p(y|x)^{\lambda}\right)^{n\rho}. \end{aligned}$$

From (13), it follows that

$$\begin{aligned} & \Pr(\text{ error } \mid d_{\min} > D_n,\ d_{\max} \le n) \\ & \overset{(a)}{\le} q^{-n\rho(1-R)}\ \big(\alpha(\mathcal{C}, D_n)\big)^{\rho}\beta(\rho, D_n) \\ & \quad \cdot \left(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\right)^{n(1-\rho)}\left(\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}} p(y|0)\left(\frac{p(y|x)}{p(y|0)}\right)^{\lambda}\right)^{n\rho} \\ & \overset{(b)}{\le} q^{-n\rho(1-R)}\ \big(\alpha(\mathcal{C}, D_n)\big)^{\rho}\beta(\rho, D_n) \\ & \quad \cdot \left(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\right)^{n(1-\rho)}\left(q\sum_{y\in\mathcal{Y}}\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}\right)^{n\rho} \\ & \overset{(c)}{=} q^{n\rho R}\ \big(\alpha(\mathcal{C}, D_n)\big)^{\rho}\beta(\rho, D_n)\left(\sum_{y\in\mathcal{Y}}\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}\right)^{n} \\ & = \beta(\rho, D_n)q^{-n\left\{-\log_q\left(\sum_{y\in\mathcal{Y}}\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}\right) - \rho\left(R + \frac{1}{n}\log_q \alpha(\mathcal{C}, D_n)\right)\right\}} \\ & = \beta(\rho, D_n)q^{-nE_{\mathrm{r}}\left(R + \frac{\log_q \alpha(\mathcal{C}, D_n)}{n}\right)} \end{aligned} \tag{47}$$

where (a) follows from (45) and (46), (b) follows from (43), and (c) follows from Lemma 5 with $\xi = t = 1$. Finally, (11) follows from (6) and (47).

APPENDIX D
PROOF OF THEOREM 3

Under the conditions in Theorem 3, we get from (8), (16), and (18) that

$$\begin{aligned}
&\Pr(\text{ error } \mid d_{\min} > D_n) \\
&\le \left(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\right)^{n(1-\rho)} \cdot \left[\sum_{n-t_0>D_n} \frac{P(n-t_0)}{1-\epsilon_n}\binom{n}{t_0}\left(\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}}p(y|0)\right)^{t_0}\right. \\
&\qquad \left.\sum_{\substack{t_1,\ldots,t_{q-1} \\ t_1+\cdots+t_{q-1}=n-t_0}} \binom{n-t_0}{t_1,\ldots,t_{q-1}} \prod_{x\in\mathcal{X}\setminus\{0\}}\left(\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}}p(y|0)\left(\frac{p(y|x)}{p(y|0)}\right)^{\lambda}\right)^{t_x}\right]^{\rho} \\
&= \left(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\right)^{n(1-\rho)} \cdot \left[\sum_{n-t_0>D_n} \frac{P(n-t_0)}{1-\epsilon_n}\binom{n}{t_0}\left(\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}}p(y|0)\right)^{t_0}\right. \\
&\qquad \left.\cdot\left(\sum_{x\in\mathcal{X}\setminus\{0\}}\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}}p(y|0)\left(\frac{p(y|x)}{p(y|0)}\right)^{\lambda}\right)^{n-t_0}\right]^{\rho}.
\end{aligned}$$

Next, setting λ and $g(y)$, as defined in (14), results in

$$\begin{aligned}
&\Pr(\text{ error } \mid d_{\min} > D_n) \\
&\overset{(a)}{\le} \left(\sum_{y\in\mathcal{Y}}\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}\right)^{n(1-\rho)} \cdot \left[\sum_{n-t_0>D_n} \frac{P(n-t_0)}{1-\epsilon_n}\binom{n}{t_0}\left(\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}}p(y|0)\right)^{t_0}\right. \\
&\qquad \left.\cdot\left(\sum_{x\in\mathcal{X}\setminus\{0\}}\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}}p(y|0)\left(\frac{p(y|x)}{p(y|0)}\right)^{\lambda}\right)^{n-t_0}\right]^{\rho} \\
&\overset{(b)}{=} \left(\sum_{y\in\mathcal{Y}}\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}\right)^{n(1-\rho)} \\
&\qquad \cdot\left\{\sum_{n-t_0>D_n} \frac{P(n-t_0)}{1-\epsilon_n}\binom{n}{t_0}\left[\sum_{y\in\mathcal{Y}}\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{\rho-1}\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{2}{1+\rho}}\right)\right]^{t_0}\right. \\
&\qquad \left.\cdot\left(\sum_{x\in\mathcal{X}\setminus\{0\}}\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}}p(y|0)\left(\frac{p(y|x)}{p(y|0)}\right)^{\lambda}\right)^{n-t_0}\right\}^{\rho} \\
&\overset{(c)}{=} \left(\sum_{y\in\mathcal{Y}}\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}\right)^{n(1-\rho)} \\
&\qquad \cdot\left(\sum_{n-t_0>D_n} \frac{P(n-t_0)}{1-\epsilon_n}\binom{n}{t_0}\left[\sum_{y\in\mathcal{Y}}\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{\rho-1}\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{2}{1+\rho}}\right)\right]^{t_0}\right. \\
&\qquad \cdot\left[q\sum_{y\in\mathcal{Y}}\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}\right. \\
&\qquad \left.\left.-\sum_{y\in\mathcal{Y}}\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{\rho-1}\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{2}{1+\rho}}\right)\right]^{n-t_0}\right)^{\rho}
\end{aligned} \tag{48}$$

where (a) follows from Lemma 5 with $\xi = 1$ and $t = 1$, (b) follows from Lemma 5 with $\xi = 1 - \frac{1}{\rho}$ and $t = 1$, and (c) follows from (44). The proof follows from Lemma 2 and (48).

APPENDIX E
PROOF OF THEOREM 4

Using the DS2 bound for the case at hand, it follows that

$$\begin{aligned}
&P(\text{ error } | d_{\min} > D_n) \\
&= \mathsf{E}\left[\iint_{(\mathbf{y},\mathbf{a}):p_{\mathbf{x}}(\mathbf{y},\mathbf{a}) \geq p_0(\mathbf{y},\mathbf{a}) \text{ for some } \mathbf{x}\neq\mathbf{0}\in\mathcal{C}} p(\mathbf{y},\mathbf{a}|\mathbf{0}) \; d\mathbf{y}\, d\mathbf{a} \,\Big|\, d_{\min} > D_n \right] \\
&\leq \mathsf{E}\left[\iint_{\mathbf{y},\mathbf{a}} p(\mathbf{y},\mathbf{a}|\mathbf{0}) \sum_{j=1}^{J} \left(\sum_{\mathbf{t}\in\mathcal{T}_j^*} \sum_{\mathbf{x}\in\mathcal{C}_\mathbf{t}} \left(\frac{p(\mathbf{y},\mathbf{a}|\mathbf{x})}{p(\mathbf{y},\mathbf{a}|\mathbf{0})} \right)^{\lambda_j} \right)^{\rho_j} d\mathbf{y}\, d\mathbf{a} \;\Big|\; d_{\min} > D_n \right] \\
&= \sum_{j=1}^{J} \mathsf{E}\Bigg[\iint_{\mathbf{y},\mathbf{a}} \psi_j(\mathbf{y},\mathbf{a}) \\
&\qquad \left(\sum_{\mathbf{t}\in\mathcal{T}_j^*} \sum_{\mathbf{x}\in\mathcal{C}_\mathbf{t}} \psi_j(\mathbf{y},\mathbf{a})^{-\frac{1}{\rho_j}} p(\mathbf{y},\mathbf{a}|\mathbf{0})^{\frac{1-\lambda_j\rho_j}{\rho_j}} p(\mathbf{y},\mathbf{a}|\mathbf{x})^{\lambda_j} \right)^{\rho_j} d\mathbf{y}\, d\mathbf{a} \,\Big|\, d_{\min} > D_n \Bigg] \qquad (49)
\end{aligned}$$

where the statistical expectation is taken over all the codebooks whose Hamming minimum distance is larger than D_n. From (49), using Jensen's inequality we have

$$\begin{aligned}
&P(\text{ error } | d_{\min} > D_n) \\
&\leq \sum_{j=1}^{J} \mathsf{E}\left[\left(\sum_{\mathbf{t}\in\mathcal{T}_j^*} \sum_{\mathbf{x}\in\mathcal{C}_\mathbf{t}} \iint_{\mathbf{y},\mathbf{a}} \psi_j(\mathbf{y},\mathbf{a})^{1-\frac{1}{\rho_j}} p(\mathbf{y},\mathbf{a}|\mathbf{0})^{\frac{1-\lambda_j\rho_j}{\rho_j}} p(\mathbf{y},\mathbf{a}|\mathbf{x})^{\lambda_j} \right)^{\rho_j} d\mathbf{y}\, d\mathbf{a} \,\Big|\, d_{\min} > D_n \right].
\end{aligned}$$

Setting $\psi_j(\mathbf{y},\mathbf{a}) = \prod_i \psi_j(y_i, a_i)$, since the channel is memoryless we have

$$\begin{aligned}
&P(\text{ error } | d_{\min} > D_n) \\
&\leq \sum_{j=1}^{J} \mathsf{E}\left[\left(\sum_{\mathbf{t}\in\mathcal{T}_j^*} \sum_{\mathbf{x}\in\mathcal{C}_\mathbf{t}} \iint_{\mathbf{y},\mathbf{a}} \prod_{i=1}^{n} \psi_j(y_i,a_i)^{-\frac{1}{\rho_j}} p_0(y_i,a_i)^{\frac{1-\lambda_j\rho_j}{\rho_j}} p_{x_i}(y_i,a_i)^{\lambda_j}\, dy_i\, da_i \right)^{\rho_j} \,\Big|\, d_{\min} > D_n \right] \\
&= \sum_{j=1}^{J} \mathsf{E}\left[\left(\sum_{\mathbf{t}\in\mathcal{T}_j^*} |\mathcal{C}_\mathbf{t}| \prod_{x\in\mathcal{X}} \left(\iint_{y,a} \psi_j(y,a)^{1-\frac{1}{\rho_j}} p_0(y,a)^{\frac{1-\lambda_j\rho_j}{\rho_j}} p_x(y,a)^{\lambda_j}\, dy\, da \right)^{t_x} \right)^{\rho_j} \,\Big|\, d_{\min} > D_n \right].
\end{aligned}$$

The proof is concluded by using Jensen's inequality (for the statistical expectation) and Lemma 2.

APPENDIX F
A CLOSED-FORM EXPRESSIONS FOR THE INTEGRAL IN THEOREM 4 WHEN APPLIED TO EXAMPLE 5

Similar to [9], we will pursue a closed-form expression by examine the exponential tilting probability measure ψ in (31). Note that the joint pdf $p_x(y,a)$ to receive the noisy observation $y \in \mathbb{C}$ with a fading sample $a \geq 0$, given that the transmitted symbol is $x \in \mathcal{X}$, is given according to

$$p(y,a|x) = \frac{1}{2\pi} \exp\left(-\frac{1}{2} \big| y - a\mu(x) \big|^2 \right) p(a),$$

where

$$p(a) = 2(1+K)a \exp\Big(-(1+K)a^2 - K \Big) I_0\left(2a\sqrt{K(K+1)} \right), \quad a \geq 0,$$

is the pdf of the Rician fading sample $a \in \mathcal{A}$ with a parameter K, and $\mu(x) \triangleq \sqrt{\frac{2E_{\rm s}}{N_0}} \exp\left(j\frac{2\pi}{q}x\right)$ is the q-ary PSK modulation mapping applied in the considered scheme. In addition, ψ in (31) is easily verified to be a probability measure, as follows from:

$$\begin{aligned}
&\int_{a=0}^{\infty}\int_{y\in\mathbb{C}} p(a)\exp\left(-\frac{\alpha}{2}\left|y-au\sqrt{\frac{2E_{\rm s}}{N_0}}\right|^2-\frac{\alpha v^2a^2E_{\rm s}}{N_0}\right)dy\,da\\
&=\int_{a=0}^{\infty}\left(p(a)\exp\left(-\frac{\alpha v^2a^2E_{\rm s}}{N_0}\right)\int_{y_{\rm r}=-\infty}^{\infty}\exp\left(-\frac{\alpha}{2}\left(y_{\rm r}-au_{\rm r}\sqrt{\frac{2E_{\rm s}}{N_0}}\right)^2\right)dy_{\rm r}\right.\\
&\qquad\left.\int_{y_{\rm i}=-\infty}^{\infty}\exp\left(-\frac{\alpha}{2}\left(y_{\rm i}-au_{\rm i}\sqrt{\frac{2E_{\rm s}}{N_0}}\right)^2\right)dy_{\rm i}\right)da\\
&=\frac{2\pi}{\alpha}\int_{a=0}^{\infty}p(a)\exp\left(-\frac{\alpha v^2a^2E_{\rm s}}{N_0}\right)da
\end{aligned}$$

where the above subscripts r and i designate the real and the imaginary parts of a complex-valued number. Next, we have the following technical lemma which evaluates the integral in (29) with the proposed tilting measure in (31):

Lemma 7. Assuming ψ as in (31), then for every $x \in \mathcal{X}$

$$\begin{aligned}
&\int_{a=0}^{\infty}\int_{y\in\mathbb{C}}\psi(y,a)^{1-\frac{1}{\rho}}p_0(y,a)^{\frac{1-\lambda\rho}{\rho}}p_x(y,a)^{\lambda}\,dy\,da\\
&=\frac{\rho}{1-\alpha(1-\rho)}\left(\frac{1+K}{\alpha(1+K+\beta)}\exp\left(-\frac{\beta K}{1+K+\beta}\right)\right)^{\frac{1}{\rho}-1}\frac{1+K}{1+K+\gamma_x}\exp\left(-\frac{\gamma_xK}{1+K+\gamma_x}\right)
\end{aligned}$$

where

$$\begin{aligned}
\beta &\triangleq \frac{\alpha v^2E_{\rm s}}{N_0}\\
\gamma_x &\triangleq \beta\left(1-\frac{1}{\rho}\right)-\frac{\rho E_{\rm s}}{\big(1-\alpha(1-\rho)\big)N_0}\left|\alpha u\left(1-\frac{1}{\rho}\right)+\frac{1-\lambda\rho}{\rho}+\lambda e^{j\frac{2\pi}{q}x}\right|^2\\
&\qquad+\frac{E_{\rm s}}{N_0}\left(\alpha\left|u\right|^2\left(1-\frac{1}{\rho}\right)+\frac{1}{\rho}\right).
\end{aligned}$$

Proof: Assuming that $1+K+\beta>0$ (which is the case since $\alpha \geq 0$), the denominator of ψ in (31) equals

$$\int_0^{\infty}p(a)\exp\left(-\frac{\alpha v^2a^2E_{\rm s}}{N_0}\right)da=\frac{1+K}{1+K+\beta}\exp\left(-\frac{\beta K}{1+K+\beta}\right).$$

Based on this equality, we evaluate the following integral:

$$\begin{aligned}
&\int_{a=0}^{\infty}\int_{y\in\mathbb{C}} \psi(y,a)^{1-\frac{1}{\rho}} p_0(y,a)^{\frac{1-\lambda\rho}{\rho}} p_x(y,a)^{\lambda}\,dy\,da \\
&= \left(\frac{1+K}{1+K+\beta}\exp\left(-\frac{\beta K}{1+K+\beta}\right)\right)^{\frac{1}{\rho}-1} \\
&\quad \int_{a=0}^{\infty}\int_{y\in\mathbb{C}} \left(\frac{\alpha p(a)}{2\pi}\right)^{1-\frac{1}{\rho}} \exp\left(-\left(1-\frac{1}{\rho}\right)\frac{\alpha}{2}\left|y-au\sqrt{\frac{2E_{\mathrm{s}}}{N_0}}\right|^2-\left(1-\frac{1}{\rho}\right)\frac{\alpha v^2a^2E_{\mathrm{s}}}{N_0}\right) \\
&\quad \left(\frac{1}{2\pi}\right)^{\frac{1-\lambda\rho}{\rho}} \exp\left(-\frac{1-\lambda\rho}{2\rho}\left|y-a\mu(0)\right|^2\right) p(a)^{\frac{1-\lambda\rho}{\rho}} \\
&\quad \left(\frac{1}{2\pi}\right)^{\lambda} \exp\left(-\frac{\lambda}{2}\left|y-a\mu(x)\right|^2\right) p(a)^{\lambda}\,dy\,da \\
&= \frac{1}{2\pi}\left(\frac{1}{\alpha}\frac{1+K}{1+K+\beta}\exp\left(-\frac{\beta K}{1+K+\beta}\right)\right)^{\frac{1}{\rho}-1} \\
&\quad \int_{a=0}^{\infty}\int_{y\in\mathbb{C}} p(a)\exp\left(-\frac{\alpha\left(1-\frac{1}{\rho}\right)}{2}\left|y-au\sqrt{\frac{2E_{\mathrm{s}}}{N_0}}\right|^2-\frac{\alpha\left(1-\frac{1}{\rho}\right)v^2a^2E_{\mathrm{s}}}{N_0}\right. \\
&\qquad\qquad \left.-\frac{1-\lambda\rho}{2\rho}\left|y-a\mu(0)\right|^2-\frac{\lambda}{2}\left|y-a\mu(x)\right|^2\right)dy\,da \\
&= \frac{1}{2\pi}\left(\frac{1}{\alpha}\frac{1+K}{1+K+\beta}\exp\left(-\frac{\beta K}{1+K+\beta}\right)\right)^{\frac{1}{\rho}-1} \\
&\quad \int_{a=0}^{\infty} p(a)\exp\left(-\frac{\alpha\left(1-\frac{1}{\rho}\right)v^2a^2E_{\mathrm{s}}}{N_0}\right) \\
&\qquad \int_{y_{\mathrm{r}}=-\infty}^{\infty}\exp\left(-\frac{\alpha\left(1-\frac{1}{\rho}\right)}{2}\left(y_{\mathrm{r}}-au_{\mathrm{r}}\sqrt{\frac{2E_{\mathrm{s}}}{N_0}}\right)^2\right. \\
&\qquad\qquad \left.-\frac{1-\lambda\rho}{2\rho}\Big(y_{\mathrm{r}}-a\mu_{\mathrm{r}}(0)\Big)^2-\frac{\lambda}{2}\Big(y_{\mathrm{r}}-a\mu_{\mathrm{r}}(x)\Big)^2\right)dy_{\mathrm{r}} \\
&\qquad \int_{y_{\mathrm{i}}=-\infty}^{\infty}\exp\left(-\frac{\alpha\left(1-\frac{1}{\rho}\right)}{2}\left(y_{\mathrm{i}}-au_{\mathrm{i}}\sqrt{\frac{2E_{\mathrm{s}}}{N_0}}\right)^2\right. \\
&\qquad\qquad \left.-\frac{1-\lambda\rho}{2\rho}\Big(y_{\mathrm{i}}-a\mu_{\mathrm{i}}(0)\Big)^2-\frac{\lambda}{2}\Big(y_{\mathrm{i}}-a\mu_{\mathrm{i}}(x)\Big)^2\right)dy_{\mathrm{i}}\,da. \qquad (50)
\end{aligned}$$

Let

$$\nu_x \triangleq -\frac{\alpha E_{\mathrm{s}}\left(1-\frac{1}{\rho}\right)}{N_0}(u_{\mathrm{r}})^2-\frac{1-\lambda\rho}{2\rho}\big(\mu_{\mathrm{r}}(0)\big)^2-\frac{\lambda}{2}\big(\mu_{\mathrm{r}}(x)\big)^2,\quad x\in\mathcal{X} \qquad (51)$$

$$\tau_x \triangleq -\frac{\alpha E_{\mathrm{s}}\left(1-\frac{1}{\rho}\right)}{N_0}(u_{\mathrm{i}})^2-\frac{1-\lambda\rho}{2\rho}\big(\mu_{\mathrm{i}}(0)\big)^2-\frac{\lambda}{2}\big(\mu_{\mathrm{i}}(x)\big)^2,\quad x\in\mathcal{X} \qquad (52)$$

$$\zeta_x \triangleq \left(1-\frac{1}{\rho}\right)\alpha u\sqrt{\frac{2E_{\mathrm{s}}}{N_0}}+\left(\frac{1-\lambda\rho}{\rho}\right)\mu(0)+\lambda\mu(x),\quad x\in\mathcal{X} \qquad (53)$$

assuming that $1-\alpha(1-\rho)>0$, it follows that

$$\begin{aligned}
&\int_{y_{\mathrm{r}}=-\infty}^{\infty}\exp\Bigg(-\frac{\alpha\left(1-\frac{1}{\rho}\right)}{2}\left(y_{\mathrm{r}}-au_{\mathrm{r}}\sqrt{\frac{2E_{\mathrm{s}}}{N_0}}\right)^2\\
&\qquad\qquad -\frac{1-\lambda\rho}{2\rho}\Big(y_{\mathrm{r}}-a\mu_{\mathrm{r}}(0)\Big)^2-\frac{\lambda}{2}\Big(y_{\mathrm{r}}-a\mu_{\mathrm{r}}(x)\Big)^2\Bigg)dy_{\mathrm{r}}\\
&= \int_{y_{\mathrm{r}}=-\infty}^{\infty}\exp\left(-\frac{1-\alpha\left(1-\rho\right)}{2\rho}y_{\mathrm{r}}^2+a\zeta_{x,\mathrm{r}}y_{\mathrm{r}}+\nu_x a^2\right)dy_{\mathrm{r}}\\
&= \sqrt{\frac{2\pi\rho}{1-\alpha\left(1-\rho\right)}}\exp\left(\frac{a^2\zeta_{x,\mathrm{r}}^2+2\nu_x a^2\frac{1-\alpha(1-\rho)}{\rho}}{2\frac{1-\alpha(1-\rho)}{\rho}}\right)\\
&= \sqrt{\frac{2\pi\rho}{1-\alpha\left(1-\rho\right)}}\exp\left(a^2\,\frac{\rho\zeta_{x,\mathrm{r}}^2+2\nu_x\big(1-\alpha\left(1-\rho\right)\big)}{2-2\alpha\left(1-\rho\right)}\right)
\end{aligned}\tag{54}$$

where $\zeta_{x,\mathrm{r}}$ is the real part of ζ^x. Similarity, let $\zeta_{x,\mathrm{i}}$ be the imaginary part of ζ^x, it follows that

$$\begin{aligned}
&\int_{y_{\mathrm{i}}=-\infty}^{\infty}\exp\Bigg(-\frac{\alpha\left(1-\frac{1}{\rho}\right)}{2}\left(y_{\mathrm{i}}-au_{\mathrm{i}}\sqrt{\frac{2E_{\mathrm{s}}}{N_0}}\right)^2\\
&\qquad\qquad -\frac{1-\lambda\rho}{2\rho}\Big(y_{\mathrm{i}}-a\mu_{\mathrm{i}}(0)\Big)^2-\frac{\lambda}{2}\Big(y_{\mathrm{i}}-a\mu_{\mathrm{i}}(x)\Big)^2\Bigg)dy_{\mathrm{i}}\\
&= \sqrt{\frac{2\pi\rho}{1-\alpha\left(1-\rho\right)}}\exp\left(a^2\,\frac{\rho\zeta_{x,\mathrm{i}}^2+2\tau_x\big(1-\alpha\left(1-\rho\right)\big)}{2-2\alpha\left(1-\rho\right)}\right).
\end{aligned}\tag{55}$$

From (50)-(55), it follows for the chosen modulation mapping μ that

$$\begin{aligned}
&\int_{a=0}^{\infty}\int_{y\in\mathbb{C}}\psi(y,a)^{1-\frac{1}{\rho}}p_0(y,a)^{\frac{1-\lambda\rho}{\rho}}p_x(y,a)^{\lambda}\,dy\,da\\
&= \frac{\rho}{1-\alpha\left(1-\rho\right)}\left(\frac{1+K}{\alpha\left(1+K+\beta\right)}\exp\left(-\frac{\beta K}{1+K+\beta}\right)\right)^{\frac{1}{\rho}-1}\\
&\quad \int_{a=0}^{\infty}p(a)\exp\left(-a^2\left[\frac{\alpha\left(1-\frac{1}{\rho}\right)v^2E_{\mathrm{s}}}{N_0}-\frac{\rho\left|\zeta_x\right|^2+2\left(\nu_x+\tau_x\right)\left(1-\alpha\left(1-\rho\right)\right)}{2-2\alpha\left(1-\rho\right)}\right]\right)da\\
&= \frac{\rho}{1-\alpha\left(1-\rho\right)}\left(\frac{1+K}{\alpha\left(1+K+\beta\right)}\exp\left(-\frac{\beta K}{1+K+\beta}\right)\right)^{\frac{1}{\rho}-1}\int_{a=0}^{\infty}p(a)\exp\left(-a^2\gamma_x\right)da.
\end{aligned}\tag{56}$$

Assuming that $1+K+\gamma_x>0$, this concludes the proof. ■

Note that the same concept of proof applies to every other modulation, as long as the resulting channel is symmetric and memoryless, where γ_x in Lemma 7 is replaced with

$$\gamma_x=\frac{\alpha\left(1-\frac{1}{\rho}\right)v^2E_{\mathrm{s}}}{N_0}-\frac{\rho\left|\zeta_x\right|^2}{2-2\alpha\left(1-\rho\right)}-\nu_x-\tau_x$$

where ν_x, τ_x and ζ_x are given by (51), (52) and (53), respectively.

APPENDIX G
A CLOSED-FORM EXPRESSIONS FOR THE INTEGRAL IN THEOREM 4 WHEN APPLIED TO EXAMPLE 6

The following exponential tilting measure is applied:

$$\psi(y,a) = \frac{\alpha p(a)}{2\pi}\left(1+\frac{\beta}{L}\sqrt{\frac{2E_s}{N_0}}\right)^L \exp\left(-\frac{\alpha}{2}\left|y-a\sqrt{\frac{2E_\mathrm{s}}{N_0}}u\right|^2 - \beta a^2\sqrt{\frac{2E_\mathrm{s}}{N_0}}\right) \tag{57}$$

where y is complex-valued, $a,\alpha,\beta \geq 0$, are real-valued parameters, u is a complex-valued parameter, and $p(a)$ is the pdf of the fading, given in (32). The following technical lemma evaluates the integral in (29) with the proposed tilting measure:

Lemma 8. Assuming ψ as in (57), then for every $x \in \mathcal{X}$

$$\begin{aligned}
&\int_{a=0}^{\infty}\int_{y\in\mathbb{C}} \psi(y,a)^{1-\frac{1}{\rho}} p_0(y,a)^{\frac{1-\lambda\rho}{\rho}} p_x(y,a)^{\lambda}\,dy\,da \\
&= \frac{\rho\alpha^{1-\frac{1}{\rho}}L^L}{1-\alpha(1-\rho)}\left(1+\frac{\beta}{L}\sqrt{\frac{2E_\mathrm{s}}{N_0}}\right)^{L\left(1-\frac{1}{\rho}\right)} \\
&\left(L+\beta\left(1-\frac{1}{\rho}\right)\sqrt{\frac{2E_\mathrm{s}}{N_0}} + \left(1-\frac{1}{\rho}\right)\frac{\alpha|u|^2E_\mathrm{s}}{N_0} + \frac{E_\mathrm{s}}{\rho N_0}\right. \\
&\qquad \left. -\frac{\rho E_\mathrm{s}}{N_0(1-\alpha(1-\rho))}\left|\alpha u\left(1-\frac{1}{\rho}\right)+\frac{1-\lambda\rho}{\rho}+\lambda\exp\left(j\frac{2\pi x}{q}\right)\right|^2\right)^{-L}.
\end{aligned}$$

Proof: Assuming that $\alpha(1-\rho)<1$, it follows that

$$\begin{aligned}
&\int_{a=0}^{\infty}\int_{y\in\mathbb{C}} \psi(y,a)^{1-\frac{1}{\rho}} p_0(y,a)^{\frac{1-\lambda\rho}{\rho}} p_x(y,a)^{\lambda}\,dy\,da \\
&= \int_{a=0}^{\infty}\int_{y\in\mathbb{C}} \left(\frac{\alpha p(a)}{2\pi}\left(1+\frac{\beta}{L}\sqrt{\frac{2E_\mathrm{s}}{N_0}}\right)^L\right)^{1-\frac{1}{\rho}} \exp\left(-\frac{\alpha}{2}\left(1-\frac{1}{\rho}\right)\left|y-au\sqrt{\frac{2E_\mathrm{s}}{N_0}}\right|^2 - \beta a^2\left(1-\frac{1}{\rho}\right)\sqrt{\frac{2E_\mathrm{s}}{N_0}}\right) \\
&\qquad \left(\frac{1}{2\pi}\right)^{\frac{1-\lambda\rho}{\rho}} \exp\left(-\frac{1-\lambda\rho}{2\rho}\left|y-a\sqrt{\frac{2E_\mathrm{s}}{N_0}}\right|^2\right) p(a)^{\frac{1-\lambda\rho}{\rho}} \\
&\qquad \left(\frac{1}{2\pi}\right)^{\lambda} \exp\left(-\frac{\lambda}{2}\left|y-a\sqrt{\frac{2E_\mathrm{s}}{N_0}}\exp\left(j\frac{2\pi}{q}x\right)\right|^2\right) p(a)^{\lambda}\,dy\,da \\
&= \frac{1}{2\pi}\left(\alpha\left(1+\frac{\beta}{L}\sqrt{\frac{2E_\mathrm{s}}{N_0}}\right)^L\right)^{1-\frac{1}{\rho}} \int_{a=0}^{\infty} p(a)\exp\left(-\beta a^2\left(1-\frac{1}{\rho}\right)\sqrt{\frac{2E_\mathrm{s}}{N_0}}\right) \\
&\int_{y\in\mathbb{C}} \exp\left(-\frac{\alpha}{2}\left(1-\frac{1}{\rho}\right)\left|y-au\sqrt{\frac{2E_\mathrm{s}}{N_0}}\right|^2 - \frac{1-\lambda\rho}{2\rho}\left|y-a\sqrt{\frac{2E_\mathrm{s}}{N_0}}\right|^2 - \frac{\lambda}{2}\left|y-a\sqrt{\frac{2E_\mathrm{s}}{N_0}}\exp\left(j\frac{2\pi}{q}x\right)\right|^2\right) dy\,da
\end{aligned} \tag{58}$$

$$\stackrel{(a)}{=} \frac{2\pi\rho}{1-\alpha\left(1-\rho\right)}\left(\alpha\left(1+\frac{\beta}{L}\sqrt{\frac{2E_{\mathrm{s}}}{N_0}}\right)^{L}\right)^{1-\frac{1}{\rho}}\frac{L^L}{\pi\left(L-1\right)!}$$
$$\int_{a=0}^{\infty} a^{2L-1}\exp\left(-a^2\left(L+\beta\left(1-\frac{1}{\rho}\right)\sqrt{\frac{2E_{\mathrm{s}}}{N_0}}-\frac{\rho\left|\zeta_x\right|^2+2\left(\nu_x+\tau_x\right)\left(1-\alpha\left(1-\rho\right)\right)}{2-2\alpha\left(1-\rho\right)}\right)\right)da$$
$$\stackrel{(b)}{=} \frac{L^L\rho}{1-\alpha\left(1-\rho\right)}\left(\alpha\left(1+\frac{\beta}{L}\sqrt{\frac{2E_{\mathrm{s}}}{N_0}}\right)^{L}\right)^{1-\frac{1}{\rho}}$$
$$\left(L+\beta\left(1-\frac{1}{\rho}\right)\sqrt{\frac{2E_{\mathrm{s}}}{N_0}}-\frac{\rho\left|\zeta_x\right|^2+2\left(\nu_x+\tau_x\right)\left(1-\alpha\left(1-\rho\right)\right)}{2-2\alpha\left(1-\rho\right)}\right)^{-L} \tag{59}$$

where (a) follows from (54) and (55), ν_x, τ_x, and ζ_x are as defined in (51)-(53), and (b) follows under the assumption that

$$0 \;<\; L+\beta\left(1-\frac{1}{\rho}\right)\sqrt{\frac{2E_{\mathrm{s}}}{N_0}}-\frac{\rho E_{\mathrm{s}}}{N_0\left(1-\alpha\left(1-\rho\right)\right)}\cdot\left|\alpha u\left(1-\frac{1}{\rho}\right)+\frac{1-\lambda\rho}{\rho}+\lambda\exp\left(j\frac{2\pi x}{q}\right)\right|^2$$
$$+\frac{E_{\mathrm{s}}}{N_0}\left(\alpha\left|u\right|^2\left(1-\frac{1}{\rho}\right)+\frac{1}{\rho}\right).$$

■

REFERENCES

[1] 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, June 2006.

[2] 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.

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

[5] S. Shamai and I. Sason, "Variations on the Gallager bounds, connections and applications," *IEEE Trans. on Information Theory,* vol 48, pp. 3029–3051, December 2002.

[6] D. Divsalar, "A simple tight bound on error probability of block codes with application to turbo codes," the Telecommunications and Mission Operations (TMO) Progress Report 42–139, JPL, pp. 1–35, November 15, 1999. [Online]. Available: http://tmo.jpl.nasa.gov/progress_report/42-139/139L.pdf.

[7] X. Wu, H. Xiang, and C. Ling, "New Gallager bounds in block-fading channels," *IEEE Trans. on Information Theory*, vol. 53, no. 2, pp. 684–694, September 2001.

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

[9] I. Sason, S. Shamai, and D. Divsalar, "Tight exponential upper bounds on the ML decoding error probability of block codes over fully-interleaved fading channels," *IEEE Trans. on Communications*, vol. 51, no. 8, pp. 1296–1305, August 2003.

[10] N. Shulman, and M. Feder, "Random coding techniques for nonrandom codes," *IEEE Trans. on Information Theory*, vol. 45, pp. 2101–2104, September 1999.

[11] A. Bennatan and D. Burshtein, "On the application of LDPC codes to arbitrary discrete-memoryless channels," *IEEE Trans. on Information Theory*, vol. 50, no. 3, pp. 417–438, March 2004.

[12] A. Bennatan and D. Burshtein, "Design and analysis of nonbinary LDPC Codes for arbitrary discrete memoryless channels," *IEEE Trans. on Information Theory*, vol. 52, no. 2, pp. 549–583, February 2006.

[13] R. G. Gallager, *Information Theory and Reliable Communication,* John Wiley and Sons, 1968.

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

[15] U. Erez and G. Miller, "The ML decoding performance of LDPC ensembles over $\mathbb{Z}_q$," *IEEE Trans. on Information Theory*, vol. 51, no. 5, pp. 1871–1879, May 2005.
[16] R. Liu, P. Spasojevic, and E. Soljanin, "Reliable channel regions for good binary codes transmitted over parallel channels," *IEEE Trans. Information Theory*, vol. 52, no. 4, pp. 1405–1424, April 2006.
[17] I. Sason and I. Goldenberg, "Coding for parallel channels: Gallager bounds and applications to turbo-like codes," *IEEE Trans. on Information Theory*, vol. 53, no. 7, pp. 2394–2428, July 2007.
[18] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell System Technical Journal*, vol. 38, pp. 611–656, May 1959.
[19] C. Shannon, R. Gallager and E. Berlekamp, "Lower bounds to error probability for decoding on discrete memoryless channels," *Information and Control*, vol. 10, Part 1: pp. 65–103, and Part 2: pp. 522–552, February/May 1967.
[20] A. Valembois and M. Fossorier, "Sphere-packing bounds revisited for moderate block length," *IEEE Trans. on Information Theory*, vol. 50, no. 12, pp. 2998–3014, December 2004.
[21] G. Wiechman and I. Sason, "An improved sphere-packing bound for finite-length codes on symmetric memoryless channels," *IEEE Trans. on Information Theory*, vol. 54, no. 5, pp. 1962–1990, May 2008.
[22] R. G. Gallager, *Low-density parity-check codes*, MA, USA:MIT Press, 1963.
[23] C. C. Wang, S. R. Kulkarni, and H. V. Poor, "Finite-dimensional bounds on $\mathbb{Z}_m$ and binary LDPC codes with belied propagation decoders," *IEEE Trans. Information Theory*, vol. 53, no. 1, pp. 56–81, January 2007.
[24] V. Rathi and R. Urbanke, "Density evolution, stability condition, thresholds for non-binary LDPC codes," *IEE Communication Proceedings*, vol. 152, no. 6, pp. 1069–1074, December 2005.
[25] M.C. Davey and D.J.C. Mackay, "Low-density parity check codes over $GF(q)$," *IEEE Communications Letters*, vol. 2, no. 6, pp. 165–167, June 1998.
[26] B. Zhou, L. Zhang, J. Kang, Q. Huang, S. Lin, and M. Xu, "Non-binary LDPC codes vs. Reed-Solomon codes," *Proceedings of the 2008 Information Theory and Applications (ITA) Workshop*, San Diago, CA, January 27–February 2, 2008. [Online]. Available: http://ita.ucsd.edu/workshop/08/files/paper/paper_151.pdf.
[27] P. Robertson and T. Woerz, "Bandwidth-efficient turbo trellis-coded modulation using punctured component codes," *IEEE Journal on Selected Areas in Communicatations*, vol. 16, no. 2, pp. 206-218, February 1998.
[28] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Bandwidth efficient parallel concatenated coding schemes," *IEE Electronics Letters*, vol. 31, no. 24, pp. 2067-2069, 1995.
[29] T. Duman and M. Salehi, "Performance bounds for turbo-coded modulation systems," *IEEE Trans. Communications*, vol. 47, no. 4, pp. 511-521, April 1999.
[30] U. Wachsmann, R. F. Fischer, and J. B. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Trans. Information Theory*, vol. 45, pp. 1361-1391, July 1999.
[31] R. Liu, J. Luo, and P. Spasojevic, "Adaptive transmission with variable-rate turbo bit-interleaved coded modulation," *IEEE Trans. Wireless Communications,* vol. 6, no. 11, pp. 3926–3936, November 2007.
[32] S. Tong, "Tangential-sphere bounds on the ensemble performance of ML decoded Gallager codes via their exact ensemble distance spectrum," to be presented at the *2008 IEEE International Conference on Communications (ICC 2008)*, Beijing, China, May 19–23, 2008.
[33] A. J. Viterbi and J. K. Omura, *Principle of Digital Communication and Coding*, 1979.
[34] D. Divsalar and E. Biglieri, "Upper bounds to error probabilities of coded systems beyond the cutoff rate," *IEEE Trans. on Communications*, vol. 51, no. 12, pp. 2011–2018, December 2003.
[35] M. Twitto, I. Sason and S. Shamai, "Tightened upper bounds on the ML decoding error probability of binary linear block codes," *IEEE Trans. on Information Theory*, vol. 53, no. 4, pp. 1495–1510, April 2007.
[36] G. Miller and D. Burshtein, "Bounds on the maximum-likelihood decoding error probability of low-density-parity-check codes," *IEEE Trans. on Information Theory*, vol. 47, pp. 2696–2710, November 2001.
[37] I. Sason and S. Shamai, "Improved upper bounds on the ensemble performance of ML decoded low-density parity-check codes," *IEEE Communications Letters*, vol. 4, no. 3, pp. 89–91, March 2000.
[38] G. Poltyrev, "Bounds on the decoding error probability of binary linear codes via their spectra," *IEEE Trans. Information Theory*, vol. 40, no. 4, pp. 1284–1292, July 1994.
[39] D. E. Knuth, *The Art of Computer Programming.* Vol.2: Seminumerical Algorithms, (3rd edition), pp. 461, Addison-Wesley, 1998.