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

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 block 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 are based on the Gallager bounds and their variations, are applied to the Gallager ensembles of non-binary and regular low-density parity-check (LDPC) codes. These upper bounds are also compared with 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 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 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 the performance analysis of 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 this bounding technique, the bounds of Duman and Salehi (see [3] and [4]) possess the pleasing feature that they are amenable to analysis for codes or ensembles for which the (average) distance spectra are available.

The bounds of Duman and Salehi, in particular its second version (called hereafter the 'DS2 bound'), are 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 shown in [1] and [5]. Gallager-based bounds for binary linear block codes whose communication takes place over fading channels are

This work was presented in part at the 2008 IEEE Information Theory Workshop (ITW 2008) which took 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).

provided in [7], [8] and [9]. The Shulman and Feder bound (SFB) [10] forms an extension of the 1965 Gallager bound which can be also 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 et al. [19], forms an alternative lower bound on the decoding error probability of block codes which applies to discrete memoryless channels. An overview on 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 bounds 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 channel is memoryless binary-input output-symmetric (MBIOS). In 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. Their asymptotic performance under iterative decoding is studied in [12], and further bounds on the thresholds of non-binary LDPC code ensembles are studied in [23] and [24]. It is assumed in [11] that the transmission takes place over channels with a random coset mechanism which enables to dismiss the channel symmetry condition required in [22]. The decoding error probability of various non-binary LDPC code constructions was studied empirically in the literature, e.g., [25]. Except for non-binary LDPC codes, turbo codes were also considered for high spectral efficiency schemes (see e.g., [26]-[30] and references therein).

The drawback of the union bound motivates the 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 over memoryless symmetric channels. Our definition of symmetry for channels whose input is non-binary generalizes the common definition of MBIOS channels. Under these symmetry requirements, we prove that the conditional error probability under ML decoding is independent of the transmitted codeword. This result is in agreement with [34] and [35] which prove the same result under linear-programming decoding.

The general concept used in this paper is based on a partitioning of the original ensemble into two subsets of codebooks according to their minimal Hamming distance. 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 their distance properties. As for the complementary set of codebooks (whose minimal Hamming distance 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 code ensembles of Gallager [22], and their error performance is studied for various communication channel models. The exact complete composition spectra for these LDPC code ensembles are also provided (instead of the upper bound in [22]), and this exact analysis forms a generalization of the analysis in [31] and [32]. In addition, the derived upper bounds are compared with sphere-packing lower bounds on the decoding error probability.

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 for the Gallager LDPC code ensembles over a q-ary symmetric and AWGN channels. Variations of these bounds are also derived and exemplified in Section IV for fully-interleaved fading channels with perfect CSI at the receiver. Section V concludes our discussion. Various technical details are relegated to the appendices.

II. Channel Symmetry and Message Independence

Let $\mathcal{X} = \{x_0, x_1, \ldots, x_{q-1}\}$ be a given alphabet with 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 (Channel symmetry). 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 $x_1, x_2 \in \mathcal{X}$ and $y \in \mathcal{Y}$, the following equality holds:

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

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

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

Example 1 (MBIOS channels). For the particular case of channels with a binary-input alphabet, and whose output alphabet $\mathcal{Y}$ is the set of real numbers, setting

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

then Definition 1 coincides with the standard definition of MBIOS channels. The meaning of the function $\mathcal{T}$ is better understood via the setting of MBIOS channels. Referring to (1), the transition probability given a channel input x_1 is equal to the transition probability given another input x_2 where the sign of the output is changed if the two binary inputs are different.

Example 2 (Random coset mechanism followed by an arbitrary channel). In [11], [13] and [14], the transmission of block codes takes place over an arbitrary memoryless channel followed by a random coset mechanism. 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 of the equivalent channel is guaranteed. 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 guaranteed 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.

Based on Definition 1, we get the following lemma:

Lemma 1. let x_1, x_2, x_3 be arbitrary symbols in $\mathcal{X}$, and let p be a transition probability law of a memoryless symmetric channel. Then,

$$p\Big(\mathcal{T}\big(\mathcal{T}(y, x_1), x_2\big)|x_3\Big) = p\big(\mathcal{T}(y, x_1 + x_2)|x_3\big) \tag{2}$$

where $\mathcal{T}$ is the mapping satisfying the symmetry properties in Definition 1.

Proof: See Appendix A. ■

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 (see [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).

[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, Section III-A].

Lemma 2. 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. ■

Lemma 2 is also valid for symmetric DMCs in the sense defined by Gallager in [13, p. 94] (as shown in the following definition):

Definition 2 (Gallager's definition for symmetric DMC [13]). A DMC is defined to be symmetric if the set of outputs can be partitioned into subsets in such a way that for each subset the matrix of transition probabilities (using inputs as rows and outputs of the subset as columns) has the property that each row is a permutation of each other row and each column (if more than 1) is a permutation of each other column.

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

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

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 (for the case of linear-programming decoding, see [34]):

Proposition 1 (Independence of the Conditional Error Probability on the Transmitted Codeword for all Memoryless Symmetric Channels). Let $\mathcal{C}$ be a linear block code whose transmission takes place over a memoryless and symmetric channel according to Definition 1. Then, the block error probability under ML decoding is independent of the transmitted codeword.

Proof: See Appendix B. ■

The proof for the message independence property remains valid even if the channel transition probability is different for each transmission. This enables the analysis in Section IV of q-ary PSK systems whose transmission takes place over fading channels with perfect CSI at the transmitter. In addition, note that in contrast to Lemma 2, Proposition 1 does not necessarily hold for symmetric DMCs as in Definition 2. This is demonstrated in the following counter-example:

Example 3 (Channel symmetry according to Definition 2 doesn't imply symmetry according to Definition 1). Consider a DMC with the integer ring $\mathbb{Z}_4$ (with arithmetic operations modulo-4) as common input and output alphabets, and with the following transition probability matrix:

$$P = [p_{i,j}] = \begin{pmatrix} 0.20 & 0.24 & 0.30 & 0.26 \\ 0.30 & 0.20 & 0.26 & 0.24 \\ 0.24 & 0.26 & 0.20 & 0.30 \\ 0.26 & 0.30 & 0.24 & 0.20 \end{pmatrix}.$$

In this matrix, the element $p_{i,j}$ (where $i, j \in \{1, \ldots, 4\}$) refers to the transition probability when the channel input is equal to $i-1$ and the output is equal to $j-1$. The memoryless channel which corresponds to P is symmetric according to Definition 2 (notice that each row and column is a permutation of another row or column, respectively). However, if the linear block code $\{00, 13, 22, 31\}$ is transmitted over the considered channel, then 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. Hence, due to Proposition 1, this channel is not symmetric according to Definition 1 although it is symmetric according to Definition 2.

III. Gallager Bounds for Memoryless Symmetric Channels and Some Applications

A. The DS2 bound

Let $\mathcal{C}$ be an (n, k) linear block code defined over the input-alphabet $\mathcal{X}$ with cardinality q. Consider the conditional error probability under ML decoding given that the m-th message is transmitted, denoted by $P_{\mathrm{e}|m}$. The DS2 bound on the conditional error probability (see [1], [3], [4] and [5]) gets the form

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

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. Here $p_n(\mathbf{y}|\mathbf{x})$ designates the transition probability of the channel where $\mathbf{x} \in \mathcal{C}$ is the transmitted codeword and $\mathbf{y} \in \mathcal{Y}^n$ is the received vector. Notice that the bound in (3) holds for an arbitrary channel regardless of its input alphabet.

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

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

where $g : \mathcal{Y} \rightarrow \mathbb{R}_+$ is an arbitrary non-negative function which is defined over the set $\mathcal{Y}$. The following bound on the decoding error probability is obtained for a discrete output alphabet (a similar proposition can be stated for channels with a 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

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

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

Proof: See Appendix C. ■

B. Performance evaluation of ensembles of linear block codes

Definition 3 (Composition of a vector). Let $\mathbf{c}$ be a vector whose components are symbols in an alphabet $\mathcal{X}$ of size q. Let us assume without loss of generality that $\mathcal{X} = \{0, \ldots, q-1\}$. The composition of $\mathbf{c}$, denoted by $\mathbf{t} = \mathbf{t}(\mathbf{c})$, is a vector $\mathbf{t} = (t_0, t_1, \ldots, t_{q-1})$ where t_x (for $x \in \mathcal{X}$) counts the number of symbols in $\mathbf{c}$ that are equal to x.

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

Lemma 3. Let $\mathcal{E}$ be an ensemble of linear block codes with block length n, and let $d_{\min}$ be the random variable designating the minimum Hamming distance of a randomly selected codebook $\mathcal{C}$ from this ensemble. Assume that there exist non-negative numbers D_n and ϵ_n, such that

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

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

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

Proof:

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

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

$$\begin{aligned} \Pr(d_{\min} \leq D_n) &\leq \sum_{\{\mathbf{c}\neq\mathbf{0}:\; w_{\text{H}}(\mathbf{c})\leq D_n\}} \Pr(\mathbf{c}\in\mathcal{C}) \\ &= \sum_{\{\mathbf{t}\in\mathcal{H}:\; n-t_0\leq D_n\}} \sum_{\{\mathbf{c}:\; \mathbf{t}(\mathbf{c})=\mathbf{t}\}} \mathsf{E}\big[1_{\{\mathbf{c}\in\mathcal{C}\}}\big] \\ &= \sum_{\{\mathbf{t}\in\mathcal{H}:\; n-t_0\leq D_n\}} \mathsf{E}\big[|\mathcal{C}_{\mathbf{t}}|\big] \end{aligned} \tag{7}$$

where $1_{\{\mathbf{c}\in\mathcal{C}\}}$ denotes the indicator of the event $\{\mathbf{c} \in \mathcal{C}\}$, and the last equality follows by converting the inner summation to an expectation. ■

Later in this section, we obtain upper bounds for the first term on the RHS of (6). These bounds are expressed in terms of the composition spectrum of the considered code ensemble, and they serve to find a suitable tradeoff between the parameters D_n and ϵ_n introduced in Lemma 3. More explicitly, since these two parameters are related, one wishes to increase the parameter D_n while maintaining small values of ϵ_n. The continuation to this section relies on Lemma 3 for the derivation of some bounds, and exemplify their use to regular LDPC code ensembles.

The following theorem provides an upper bound on the decoding error probability for 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 3, the block error probability under ML decoding satisfies

$$P_{\text{e}} \leq \left(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\right)^{n(1-\rho)} \left(\sum_{\mathbf{t}\in\mathcal{H}:\; n-t_0>D_n} \mathsf{E}\Big[|\mathcal{C}_{\mathbf{t}}| \;\big|\; d_{\min}>D_n\Big] \prod_{x\in\mathcal{X}} \big(s_{\lambda,\rho}(x)\big)^{t_x}\right)^{\rho} + \epsilon_n \tag{8}$$

where

$$s_{\lambda,\rho}(x) \triangleq \sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}} p(y|0)^{1-\lambda} p(y|x)^{\lambda}, \quad x\in\mathcal{X} \tag{9}$$

and $\mathsf{E}\big[\,|\mathcal{C}_{\mathbf{t}}|\;\big|\;d_{\min}>D_n\big]$ denotes the conditional expected number of codewords whose composition is equal to $\mathbf{t}$ (where the expectation is with respect to the choice of the codebook $\mathcal{C}$ from the ensemble $\mathcal{E}$) under the requirement that the minimal Hamming weight of the randomly selected codebook is larger than D_n.

Proof: From Proposition 2 and (9), we get the following upper bounding on the first summand in (6):

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

where $\mathcal{C}_{\mathbf{t}}$ is the set of all codewords in a codebook $\mathcal{C}$ whose composition is $\mathbf{t}$. Notice that the double summations on 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 the non-zero codewords. Using Jensen's inequality, $\mathsf{E}[X^{\rho}] \leq \big(\mathsf{E}[X]\big)^{\rho}$ for $0\leq\rho\leq 1$, then

$$\begin{aligned}&\Pr(\text{ error } \mid d_{\min}>D_n)\\ \leq\;& \left(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\right)^{n(1-\rho)}\\ &\cdot\left(\sum_{\mathbf{t}\in\mathcal{H}} \mathsf{E}\left[\sum_{\mathbf{c}\in\mathcal{C}_{\mathbf{t}}}\prod_{x\in\mathcal{X}}\big(s_{\lambda,\rho}(x)\big)^{t_x} \;\middle|\; d_{\min}>D_n\right]\right)^{\rho}\\ =\;& \left(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\right)^{n(1-\rho)}\\ &\cdot\left(\sum_{\mathbf{t}\in\mathcal{H}} \mathsf{E}\Big[|\mathcal{C}_{\mathbf{t}}| \;\Big|\; d_{\min}>D_n\Big]\prod_{x\in\mathcal{X}}\big(s_{\lambda,\rho}(x)\big)^{t_x}\right)^{\rho}.\end{aligned} \tag{10}$$

For all codewords whose composition $\mathbf{t}$ satisfies $n-t_0\leq D_n$, their Hamming weight is not larger than D_n. Hence

$$\mathsf{E}\Big[|\mathcal{C}_{\mathbf{t}}| \;\Big|\; d_{\min}>D_n\Big] = 0, \quad \forall\, \mathbf{t}\in\mathcal{H}:\; n-t_0\leq D_n \tag{11}$$

and the bound in (8) follows from Lemma 3, and (10) and (11). ∎

The following theorem is a particularization of Theorem 1:

Theorem 2. Under the assumptions and notation in Proposition 2 and Lemma 3, the block error probability satisfies

$$P_{\text{e}} \leq\; q^{-nE_{\text{r}}\left(R+\frac{\log_q \alpha_q(\mathcal{C},D_n)}{n}\right)} \;+\; \epsilon_n \tag{12}$$

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

$$\begin{aligned} E_{\mathrm{r}}(R) &\triangleq \max_{0\le\rho\le 1} \left(E_0(\rho) - \rho R\right) \\ E_0(\rho) &\triangleq -\log_q \left(\sum_{y\in\mathcal{Y}} \left(\frac{1}{q} \sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}} \right)^{1+\rho} \right) \\ \alpha_q(\mathcal{C}, D_n) &\triangleq \max_{\{\mathbf{t}\in\mathcal{H}:\ n-t_0>D_n\}} \left\{ \frac{\mathsf{E}\left[|\mathcal{C}_{\mathbf{t}}| \;\middle|\; d_{\min} > D_n \right]}{q^{-n(1-R)}\binom{n}{\mathbf{t}}} \right\}. \end{aligned} \tag{13}$$

Proof: See Appendix D. ■

A similar theorem can be stated for memoryless symmetric channels with continuous-output alphabets, where sums are replaced by integrals.

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 (12) does not depend on the communication channel, but only on the code ensemble and the arbitrary threshold D_n. This partitioning differs from [11] and [37] where no such separation of codebooks is used. The SFB in [11] and [37] is combined with a union bound which corresponds to all pairwise error probabilities of relevant codewords and it depends on the communication channel. Following Example 2, the SFB in [11] can be considered 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., $\mathsf{E}\left[|\mathcal{C}_{\mathbf{t}}|\middle| d_{\min} > D_n\right]$) is not available. Nevertheless, it is possible to use the inequality

$$\begin{aligned} \mathsf{E}\Big[|\mathcal{C}_{\mathbf{t}}|\Big] &\geq \mathsf{E}\Big[|\mathcal{C}_{\mathbf{t}}| \mid d_{\min} > D_n\Big] \ \Pr(d_{\min} > D_n) \\ &\geq \mathsf{E}\Big[|\mathcal{C}_{\mathbf{t}}| \mid d_{\min} > D_n\Big] \ (1 - \epsilon_n). \end{aligned} \tag{14}$$

where the LHS of this inequality requires the knowledge of the expectation of the complete composition spectrum $\mathsf{E}\big[|\mathcal{C}_{\mathbf{t}}|\big]$. Applying (14) to the RHS of (8), gives a looser version of the bounds in Theorem 1 and 2 but is more amenable to analysis. The inequality in (14) is valid when expurgation of codebooks is considered. The expurgated ensemble is constructed by removing all codebooks whose minimum Hamming distance is not larger than D_n. Since all the 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 choose a codebook from this ensemble uniformly at random. We assume 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

$$\mathsf{E}\Big[|\mathcal{C}_{\mathbf{t}}|\Big] = P(n-t_0)\binom{n}{\mathbf{t}} \tag{15}$$

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 (15), the evaluation of α_q in Theorem 2 is considerably reduced.

In the following, we introduce an improvement over the bound in Theorem 2:

Theorem 3. Under the assumptions and notation in Proposition 2 and Lemma 3, for ensembles satisfying (15), the block error probability satisfies

$$P_{\mathsf{e}} \le A(\rho)^{n(1-\rho)} \left(\sum_{D_n < l \le n} \frac{P(l)}{1-\epsilon_n} \binom{n}{l} B(\rho)^{n-l} C(\rho)^l \right)^{\rho} + \epsilon_n \tag{16}$$

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 E. ■

Remark 2. For the particular case of binary linear block codes, the bound provided in Theorem 3 does not require the symmetry assumption on the considered ensemble in (15). For this case, the same derivation holds while setting

$$P(l) \triangleq \frac{\mathsf{E}\big[\,|\mathcal{C}_l|\,\big]}{\binom{n}{l}},\ D_n < l \leq n$$

where $\mathsf{E}\big[\,|\mathcal{C}_l|\,\big]$ denotes the expected number of codewords whose Hamming weight is l.

C. Performance of non-binary regular LDPC ensembles

The non-binary (c, d)-regular LDPC code 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 (15):

Lemma 4. 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

$$P(l) \leq \left(\frac{\exp\left(\frac{n}{d}\left(\mu_q(s) - s\mu_q'(s) + (d-1)\ln q\right)\right)}{\binom{n}{l}(q-1)^l}\right)^c \tag{17}$$

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

$$\frac{n}{d}\mu_q'(s) = l. \tag{18}$$

Note, that the bound in (17) is valid for all s, not only for the one satisfying (18) which yields the minimum bound in (17). Using the change of variables $s = \ln\frac{1-u}{1+(q-1)u}$, $-\frac{1}{q-1} \leq u \leq 1$, in (18), 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 the RHS of (17) for the binary case are provided in [8]).

In the following, we obtain the exact composition spectrum of the regular LDPC code ensembles of Gallager. This derivation serves to improve the tightness of the bounds on the error probability. The provided analysis generalize [31] to non-binary codes. The exact enumeration for the binary case is already available in [32], as an intermediate result, although its main interest is in asymptotic analysis (This analysis can be traced even to Gallager [2]).

Lemma 5. Under the assumptions and notation in Lemma 4, the probability $P(l)$ satisfies

$$P(l) = \left(\frac{A_l}{\binom{n}{l}(q-1)^l} \right)^c, \quad 2 \le l \le n \tag{19}$$

where

$$\sum_{2 \le l \le n} A_l X^l \triangleq \bigl(A^*(X)\bigr)^{\frac{n}{d}} \tag{20}$$

$$A^*(X) \triangleq 1 + \frac{1}{q} \sum_{l=2}^{d} \Bigl((q-1)^l + (q-1)(-1)^l\Bigr) \binom{d}{l} X^l. \tag{21}$$

Proof: See Appendix F. ■

As suggested in [31], the numerical evaluation of the exponent in (20) is carried out, in all the examples studied in this paper, via the binary method (see [40, p. 441]). This method makes the evaluation of the high-order powers of a polynomial relatively easy to compute.

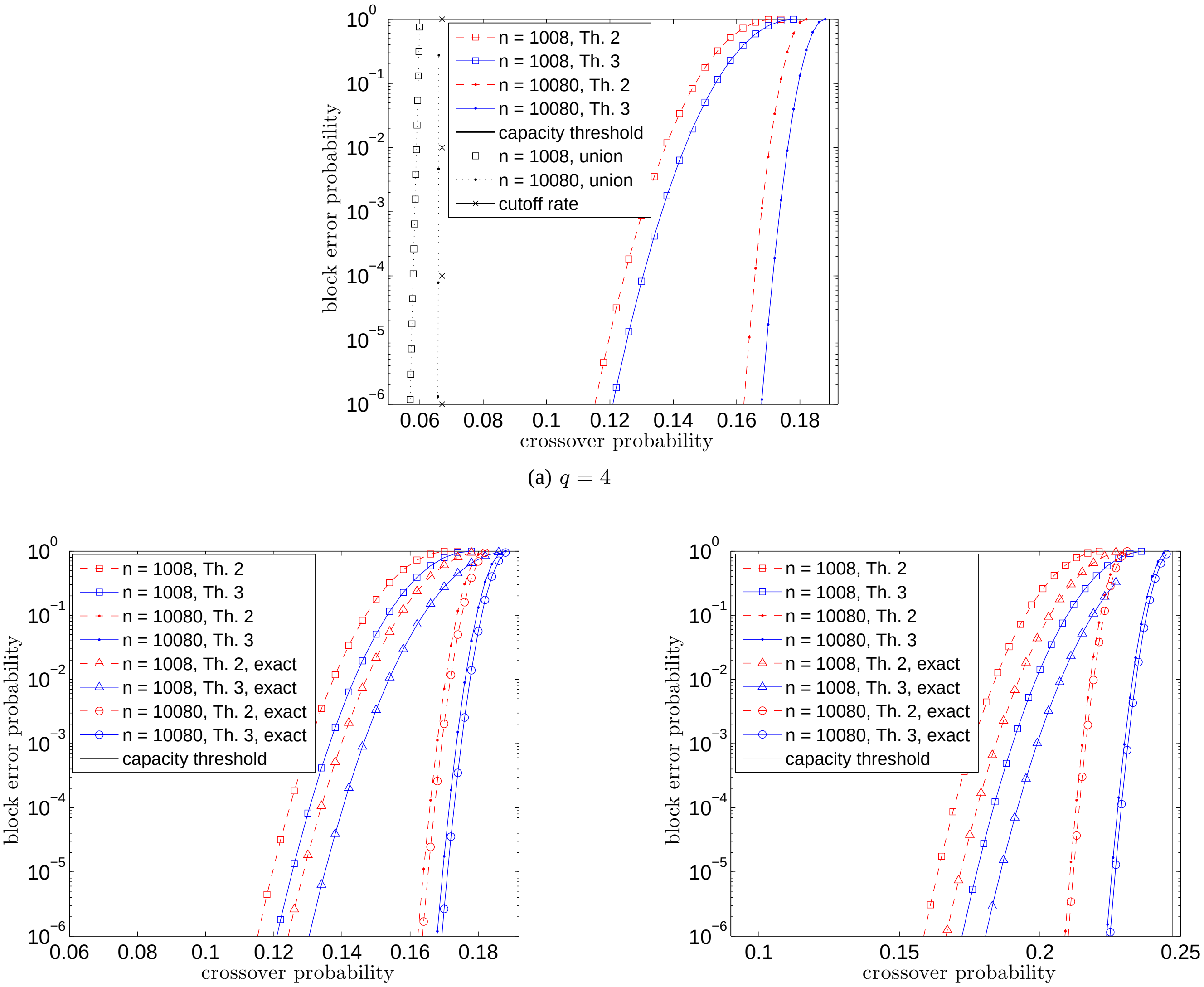
The 1961 Gallager-Fano bound (see [1], [22]) and Lemma 4 imply an exponential bound (in terms of the block length) on the decoding error probability for the expurgated LDPC code ensemble. This expurgation removes all the codebooks whose minimal Hamming distance is below a certain threshold which scales linearly with the block length. This result is elaborated for the binary case by Miller and Burshtein [37]).

The following examples consider the Gallager ensembles of non-binary and (8, 16) regular LDPC codes where these ensembles are expurgated by removing all the codebooks whose minimum distance is not greater than a certain parameter D_n. The examples study upper bounds on the decoding error probability of these expurgated ensembles via the use of the upper bounds in Theorems 2 and 3. The exact composition spectrum of the non-expurgated LDPC code ensemble is evaluated via Lemma 5, and then upper bounds on the composition spectrum of the expurgated ensembles are calculated via (14).

Example 4 (**q-ary symmetric channels).** Bounds on the block error probability for some expurgated LDPC code ensembles are presented in Figure 1 when the transmission takes place over a q-ary symmetric channel and ML decoding is performed. The performance bounds introduced in this paper are compared with the union bound, and we also exemplify the uselessness of the union bound 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 to assess the performance of efficient code ensembles which perform reliably at rates exceeding the cutoff rate of the channel. Fig. 1(b) is focused on the improved bounds in Theorems 2 and 3, applied to the Gallager $(8, 16)$ regular and expurgated LDPC code ensemble with a quaternary alphabet and block lengths of $n = 1008$ and 10080 symbols. The ensemble spectrum is upper bounded via Lemma 4, and in addition it is exactly evaluated using Lemma 5; both options are applied in this example so that the improvement provided by the exact calculation of the composition spectrum is exemplified in this figure. The various choices of the parameter D_n and the resulting ϵ_n, which serves as an upper bound on the fraction of codebooks whose minimum distance is not larger than D_n, are detailed in Table I(a). Since Theorem 3 is tighter than Theorem 2, then the minimal value of D_n for which Theorem 2 is useful is larger than the corresponding value



(a) $q = 4$

(b) $q = 4$

(c) $q = 8$

Fig. 1: Upper bounds on the block error probability of the Gallager $(8, 16)$ regular and non-binary LDPC code ensembles with quaternary and octal input alphabets. The transmission takes place over a q-ary symmetric channel where $q = 4$ in plots (a) & (b) and $q = 8$ in plot (c). This figure refers to expurgated ensembles whose block lengths are 1008 and $10,080$ symbols.

which is calculated in conjunction with Theorem 3. Moreover, the considered bounds are further improved when the upper bound for the composition spectrum in Lemma 4 is replaced with the exact calculation in Lemma 5. The inferiority of the SFB in (12) is further pronounced for higher alphabets, as exemplified for octal signaling in Figure 1(c) (where the details with regard to the choices of D_n and ϵ_n values are given in Table I(b)).

Example 5 (AWGN channels with a q-ary PSK modulation). Upper bounds on the block error probability for for some expurgated LDPC code ensembles are depicted in Figure 2 when the transmission takes place over the AWGN channel with a q-ary PSK modulation. The alphabet size of these code ensembles is $q = 4, 8, 16$, and 32, and the examined parameters D_n of the expurgation are given in Table II. It is evident that the SFB in Theorem 2 deteriorates as compared to the bound in Theorem 3. This deterioration is more dominant by increasing the alphabet size q. It is interesting to compare the studied bounds to the union bound which, for large block lengths, diverges 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 deteriorates considerably comparing to the bound provided in Theorem 3 and to the union bound which is depicted in Figure 2 and (d) (the

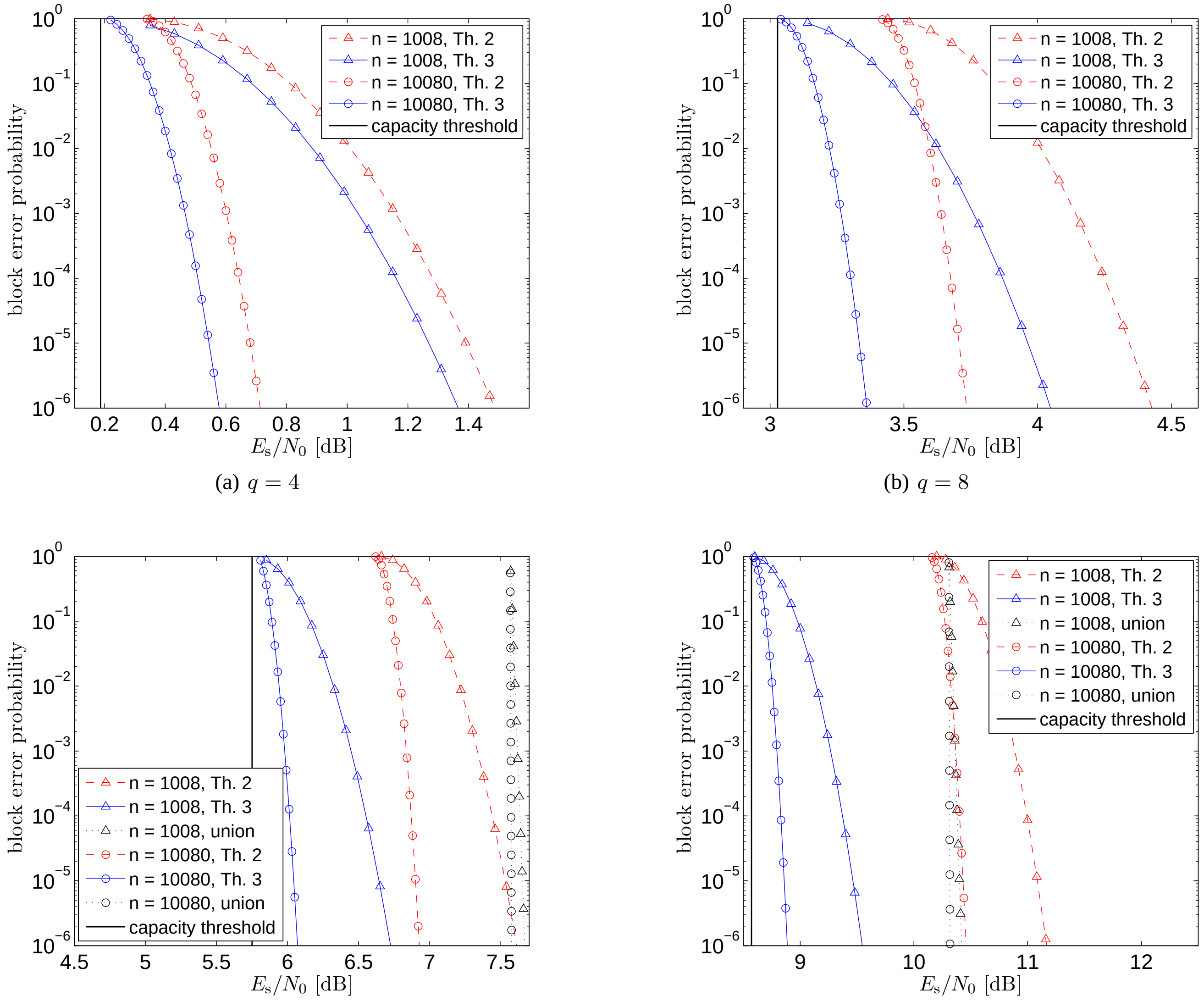
TABLE I: Parameters for Example 4

Performance bound	Block length n (symbols)	D_n	ϵ_n (Lemma 4)	ϵ_n (Lemma 5)
Theorem 2	1008	173	0.1	10^{-11}
Theorem 3	1008	99	10^{-4}	10^{-11}
Theorem 2	10008	1834	0.11	10^{-17}
Theorem 3	10008	600	10^{-7}	10^{-17}

(a) Quaternary alphabet ($q = 4$).

Performance bound	Block length n (symbols)	D_n	ϵ_n (Lemma 4)	ϵ_n (Lemma 5)
Theorem 2	1008	191	10^{-5}	10^{-14}
Theorem 3	1008	119	10^{-5}	10^{-14}
Theorem 2	10080	1951	10^{-9}	10^{-20}
Theorem 3	10080	887	10^{-9}	10^{-20}

(b) Octal alphabet ($q = 8$).



(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. This figure depicts the upper bounds on the block error probability for the expurgated ensemble with block lengths of 1008 and $10,080$ symbols.

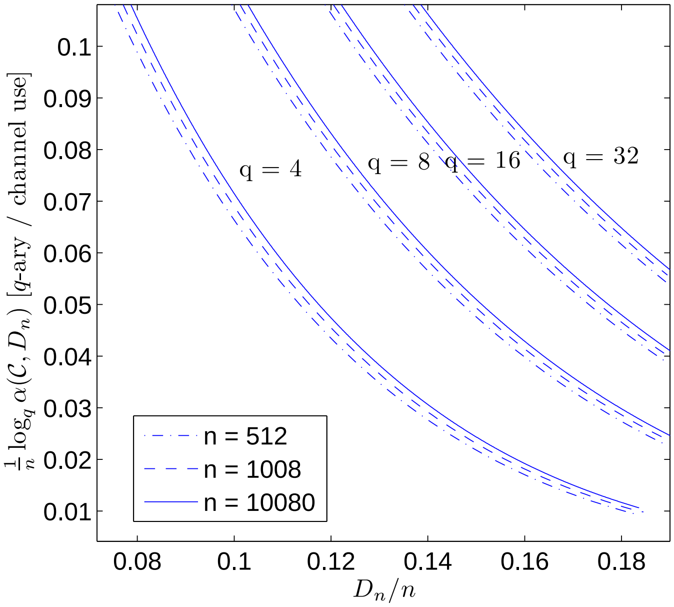


Fig. 3: The term $\frac{1}{n}\log_q \alpha_q(\mathcal{C}, D_n)$ in (12) for the regular (8,16) LDPC ensemble of Gallager [22], depicted for alphabet sizes of $q = 4$, 8, 16, and 32, and block lengths of $n = 512$, 1008, and 10080 symbols.

TABLE II: D_n values for Example 5

Performance bound	Block length n (symbols)	D_n (q=4)	D_n (q=8)	D_n (q=16)	D_n (q=32)
Theorem 2	1008	186	191	191	191
Theorem 3	1008	38	34	15	12
Theorem 2	10080	1851	1951	1951	1951
Theorem 3	10080	282	216	132	102

SNR values which correspond to the cutoff rate for $q = 16$ and 32 are equal to 7.57 dB and 10.31 dB, respectively).

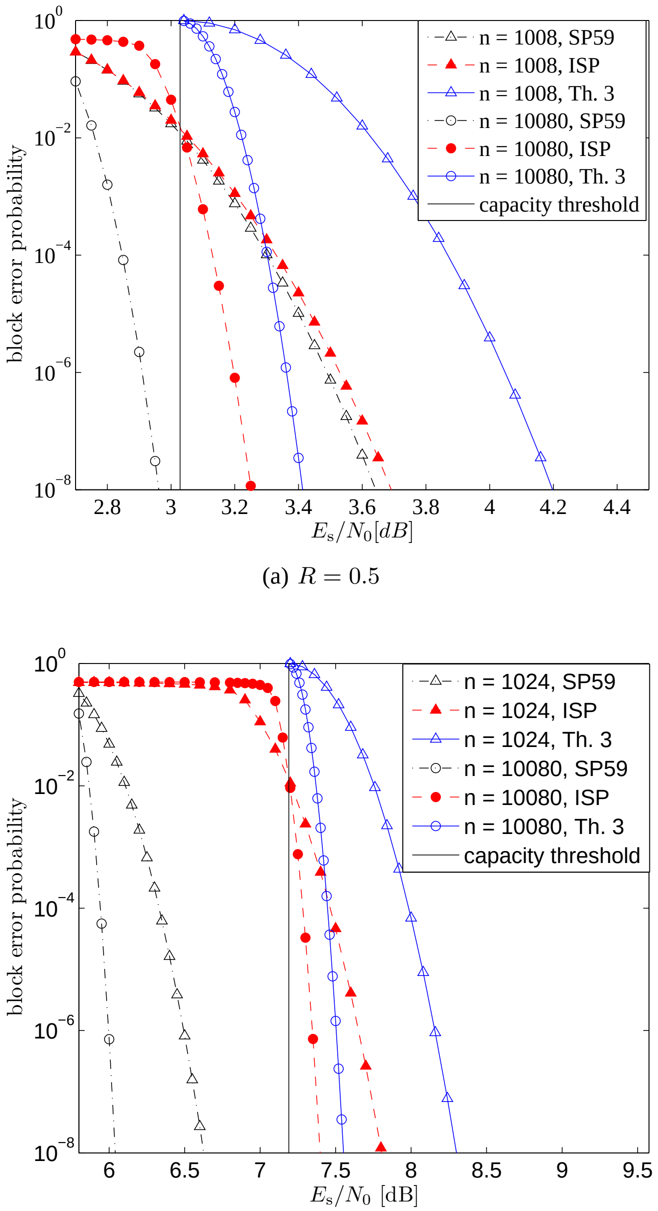
The reason for the deterioration of the SFB for large values of q is explained when looking into the rate term $\frac{1}{n}\log_q \alpha(\mathcal{C}, D_n)$. This term corresponds to the difference between the spectrum of the considered ensemble and the multinomial spectrum of the fully random code ensemble. This difference between the two composition spectra is depicted in Figure 3 as a function of $\frac{D_n}{n}$ for alphabet sizes of $q = 4$, 8, 16, and 32, and for block lengths of $n = 512$, 1008, and 10080 symbols. From Figure 3, this term is more pronounced by increasing the value of q. On the other hand, the bound in Theorem 3 does not exhibit such deterioration.

Remark 3. Divsalar's bound [6], [36] is widely used when assessing the error performance of binary turbo-like code ensembles over the binary-input AWGN channel (see [1, Chapter 3.2.4] and references therein). This is due to the fact that the bound is given in a closed form, and its calculation does not involve any numerical integrations and parameter optimizations. 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 which is centered at a point along the line connecting the origin to the all-zero codeword, and whose radius is optimized analytically in order to get the tightest bound within its form. This technique was generalized by the authors to the non-binary setup by examining various regions in the complex observation space. In contrast to the binary case, not all the parameters could be optimized analytically. Moreover, the resulting bounds were not satisfactory as compared to the bounds presented in Example 5, and are therefore omitted.

Example 6 (**A Comparison to lower bounds on the decoding error probability).** The upper bound in Theorem 3 is compared in Figure 4 to the SP59 lower bound of Shannon [18], and the ISP lower bound in [21]. The regular



(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 and $10,080$ symbols. The upper bounds are provided for expurgated $(8,16)$ and $(8,32)$ regular LDPC code ensembles.

LDPC code ensembles of Gallager are considered with octal alphabet cardinality and block lengths of 1008 and 10080 symbols, and the performance is studied over the AWGN channel with an 8-ary PSK modulation. In Fig. 4(a), the upper bound in Theorem 3 is depicted for the Gallager (8,16) regular and expurgated LDPC code ensemble with octal alphabet (the bound is evaluated with the same parameters as in Table II). 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 probability. For a block length of 1008 symbols, a negligible difference exists between the two considered lower bounds, and both of these bounds are about 0.5 dB away from the upper bound in Theorem 3 for all range of interest. For the larger block length of 10080 symbols, the gain of the ISP bound is about 0.25 dB as compared to the SP59 bound, and it is about 0.2 dB away from the upper bound (see Fig. 4(a)). The comparison between the upper and lower bounds is further studied in Fig. 4(b) for the Gallager (8,32) regular and expurgated LDPC code ensembles with block lengths of 1024 and 10080 symbols and octal alphabet. The design rate for these ensembles is 0.75 symbols per channel use. The upper bound in Theorem 3 is depicted with $D_n = 25$ and 95, respective to the studied block lengths. The ISP bound maintains its close proximity with the 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 (see Fig. 4(b)).

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 random variable (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 conditional 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, 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 4. 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 equality holds:
$$p(y|x_1, a) = p(\mathcal{T}_a(y, x_2 - x_1)|x_2, a). \tag{22}$$

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 the condition in (22) for every fading sample $a \in \mathcal{A}$. Nevertheless, this weaker notion is sufficient in order to prove that for the case at hand, the ML decoding error probability does not depend on the actual transmitted message. This is clearly expected since Definition 4 is a direct application of Definition 1 for the case at hand. The conditional decoding error probability for the m-th message under ML decoding as is given by

$$P_{\mathrm{e}|m} = \int_{\mathbf{a}} \int_{\mathbf{y} \in \Lambda_m^c(\mathbf{a})} p(\mathbf{y}, \mathbf{a}|\mathbf{x}_m)\, d\mathbf{y}\, d\mathbf{a} = \int_{\mathbf{a}} p(\mathbf{a}) \int_{\mathbf{y} \in \Lambda_m^c(\mathbf{a})} p(\mathbf{y}|\mathbf{x}_m, \mathbf{a})\, d\mathbf{y}\, d\mathbf{a} \tag{23}$$

where $\Lambda_m(\mathbf{a}) \subseteq \mathcal{Y}^n$ is the decision region under ML decoding given that the sequence of fading samples is $\mathbf{a} \in \mathcal{A}^n$. The proof of the independence of the decoding error probability on the transmitted codeword follows by showing that the inner integral in (23) 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 in Appendix B).

Theorem 4. Under the assumptions and notation in Lemma 3, consider the case where transmission takes place over a symmetric, fully-interleaved fading channel with perfect CSI at the receiver. Let the channel input and output alphabets be $\mathcal{X}$ and $\mathcal{Y}$, respectively, and let p be the transition pdf of the channel. Then, the block error probability under ML decoding satisfies

$$\begin{aligned} P_{\mathrm{e}} \;\leq\; \sum_{j=1}^{J} \Bigg(& \sum_{\mathbf{t} \in \mathcal{H}_j:\; n - t_0 > D_n} \mathsf{E}\Big[|\mathcal{C}_{\mathbf{t}}| \;\Big|\; d_{\min} > D_n\Big] \\ & \prod_{x \in \mathcal{X}} \left(\iint \psi_j(y, a)^{1 - \frac{1}{\rho_j}} p(y, a|0)^{\frac{1 - \lambda_j \rho_j}{\rho_j}} p(y, a|x)^{\lambda_j}\, dy\, da \right)^{t_x} \Bigg)^{\rho_j} \;+\; \epsilon_n \end{aligned} \tag{24}$$

where $\{\mathcal{H}_j\}_{j=1}^{J}$ with an arbitrary $J \geq 1$ forms a partition of the set of compositions (except for the one which corresponds to 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 $d_{\min} > D_n$, the functions $\psi_j : \mathcal{Y} \times \mathcal{A} \rightarrow \mathbb{R}$ are arbitrary non-negative tilting probability measures, and $0 \leq \rho_j \leq 1$ and $\lambda_j \geq 0$.

Proof: See Appendix G. ■

Consider an ensemble which satisfies the symmetry property in (15), and let us choose $J = n$ and $\mathcal{H}_j = \{\mathbf{t} : n - t_0 = j\}$. By using calculus of variations, the optimum tilting measures ψ_j for $D_n < j \leq n$, are given by

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

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

$$\alpha_{j,x} \triangleq \frac{\frac{j}{n} \iint \psi_j(y,a)^{1-\frac{1}{\rho_j}} p(y,a|0)^{\frac{1}{\rho_j}}\, dy\, da}{(1-\frac{j}{n}) \sum_{x \in \mathcal{X}^*} \iint \psi_j(y,a)^{1-\frac{1}{\rho_j}} p(y,a|0)^{\frac{1-\lambda_j \rho_j}{\rho_j}} p(y,a|x)^{\lambda_j}\, dy\, da}$$

and $\alpha_{j,0}$ are determined 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 probability tilting measures for which the integration in (24) 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 a fully-interleaved Rician fading channel with a q-ary PSK modulation.

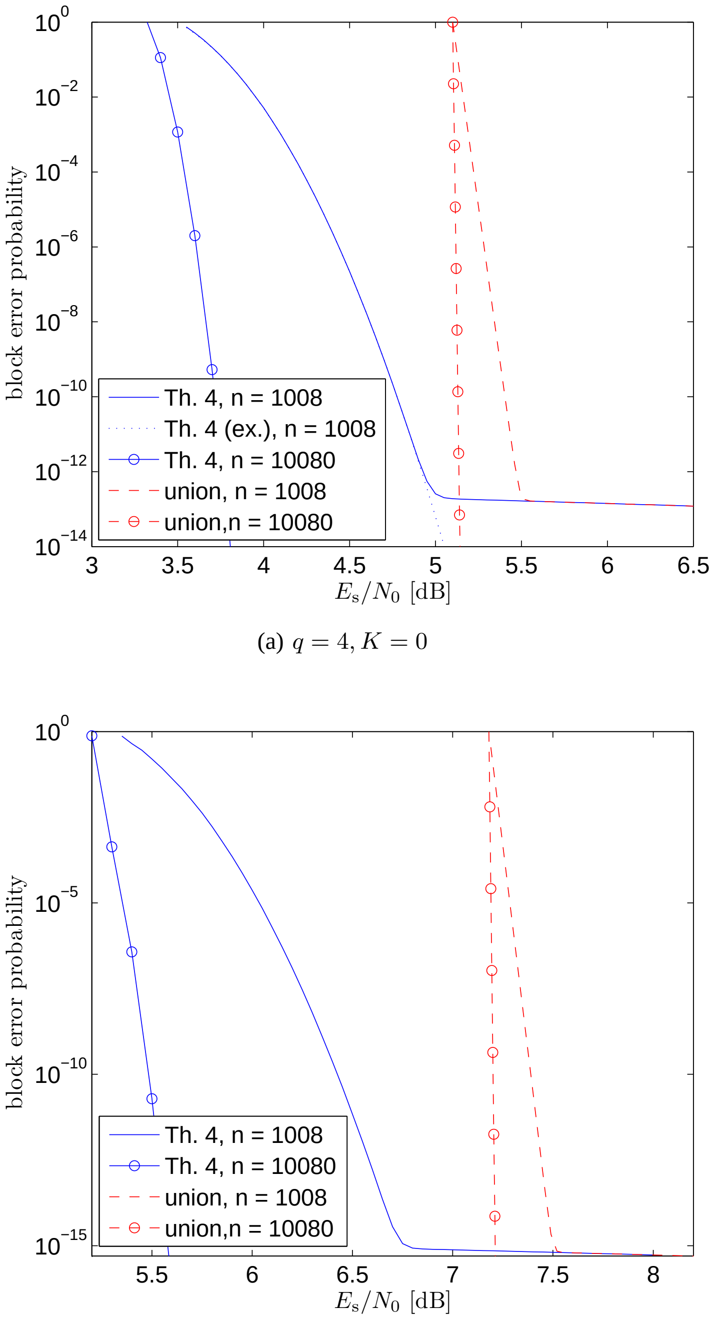
Example 7 (A fully-interleaved Rician fading channel with PSK modulation). Consider the class of fully-interleaved Rician fading channels with an additive white Gaussian noise. A codeword $\mathbf{x} = (x_1, \ldots, x_n)$ with a block length n and codeword symbols over the alphabet $\mathcal{X} = \{0, 1, \ldots, q-1\}$ is transmitted over a discrete-time memoryless channel. The received sequence $\mathbf{y} = (y_1, \ldots, y_n) \in \mathbb{C}^n$ satisfies

$$y_k = A_k \sqrt{\frac{2E_{\rm s}}{N_0}} \exp\left(\frac{2\pi i}{q} x_k\right) + N_k, \quad k = 1, \ldots, n. \tag{25}$$

Here A_k is a Rician random variable with a parameter K, and $N_k = N_k^{\rm r} + jN_k^{\rm i}$, where $N_k^{\rm r}$ and $N_k^{\rm i}$ are statistically independent Gaussian random variables 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 $E_{\rm s}$ 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

$$\psi_j(y,a) = \frac{\frac{\alpha_j}{2\pi} \exp\left(-\frac{\alpha_j}{2} \left|y - a u_j \sqrt{\frac{2E_{\rm s}}{N_0}}\right|^2 - \frac{\alpha v_j^2 a^2 E_{\rm s}}{N_0}\right) p(a)}{\int_0^\infty p(a) \exp\left(-\frac{\alpha v_j^2 a^2 E_{\rm s}}{N_0}\right) da}, \quad y \in \mathbb{C},\ a \geq 0 \tag{26}$$

where, for $1 \leq j \leq J$, v_j and α_j are non-negative real-valued parameters, and u_j is a complex-valued parameter. Substituting the exponential tilting measure ψ_j into (24) provides an upper bound on the error probability which is expressed in a closed form (see Appendix H). 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 addition to the union bound. The bound in (24) is evaluated with $J = 6$ and the partitioning of the set of compositions is done according to their Hamming weights where the boundaries of this partitioning are set to Hamming weights of 350, 425, 500, 575, and 600 for a block length of 1008 symbols (the corresponding boundaries for a block length of 10080 symbols are set to 3500, 4250, 5000, 5750, and 6000). The performance bounds refer to a quaternary input-alphabet $q = 4$ and a fully-interleaved Rayleigh fading channel (see Fig. 5(a)),



(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 q-ary PSK modulation and perfect CSI at the receiver. Both plots refer to the non-expurgated ensemble, and the performance of an expurgated ensemble with $D_n = 100$ is also presented in plot (a) for comparison.

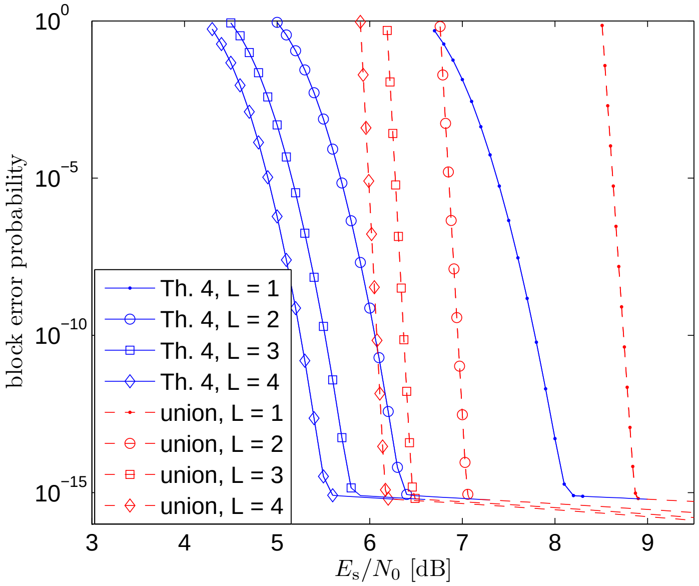


Fig. 6: Upper bounds on the block error probability under ML decoding for the $(8, 16)$-regular LDPC ensemble of Gallager with octal alphabet and a block length of 1008 symbols. The transmission takes place over a fully-interleaved Rayleigh fading channel with 8-ary PSK modulation, perfect CSI and maximal ratio combining (MRC) at the receiver. The figure depicts the performance for MRC diversity with $L = 1$ to $L = 4$ antennas at the receiver.

and for octal input-alphabet $q = 8$ and a Rician fading channel with $K = 2$ (see Fig. 5(b)). 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 $\epsilon_n = 10^{-5}$ in Theorem 4) is also presented for a block length of 1008 symbols. In both plots, the union bound diverges bellow the cutoff rate which corresponds 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 is not informative (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 Fig. 5(a) and 1.2 dB and 1.3 dB in Fig. 5(b) at block error probabilities 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 probabilities).

Example 8 (**A fully-interleaved Rayleigh fading channel with PSK modulation and maximal ratio combining).** Consider the class of fully-interleaved Rayleigh fading channels with maximal ratio combining (MRC) space diversity of order L. The receiver sequence is as in (25) where the fading samples, A_k, are distributed according to the following pdf:

$$p(a) = \frac{2L^L a^{2L-1} \exp\left(-La^2\right)}{(L-1)!}, \quad a \geq 0. \tag{27}$$

Note that $\frac{E_s}{N_0}$ in (25) refers to the stage 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 I). Consider the (8,16) regular and non-binary LDPC code ensemble of Gallager [22] with octal alphabet and a block length of 1008 symbols. Upper bounds on the decoding error probability of this ensemble with various diversity orders L are shown in Figure 6. The bound provided in Theorem 4 is compared with the union bound for MRC diversity with $L = 1$ to 4 antennas. 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 5.90 dB for $L = 1, 2, 3$ and 4 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, respectively). Nevertheless, the bound in Theorem 4 outperforms the union bound by 1.33 dB at a block error rate of 10^{-4} when there is a single antenna at the receiver, and by 1.02 dB for $L = 4$ receiving antennas.

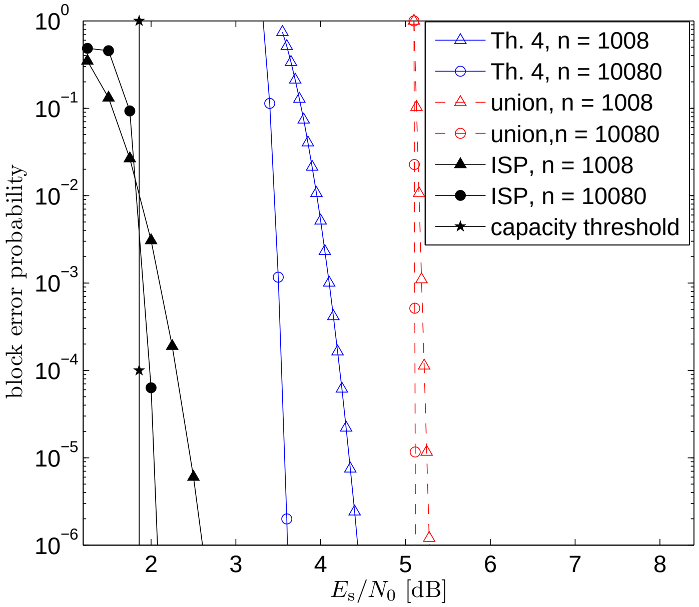


Fig. 7: A comparison between the DS2 and union upper bounds on the block error probability under ML decoding for the $(8, 16)$-regular LDPC ensemble of Gallager (see Example 7). The transmission takes place over fully-interleaved Rayleigh fading channel with a QPSK modulation and perfect CSI at the receiver. The ISP lower bounds on the decoding error probability are shown for block lengths of 1008 and 10080 symbols. The capacity limit for infinite block length is also presented as a reference.

Example 9 (**A comparison of upper and lower bounds**)**.** The DS2 upper bound in Theorem 4 is compared in this example to an improved sphere-packing (ISP) lower bound on the ultimate error performance of finite-length codes (see [21]). The bounds are compared for block codes whose transmission takes place over the fully interleaved Rayleigh fading channels with a quadrature-phase shift-keying (QPSK) modulation and perfect CSI at the receiver. The DS2 bound is evaluated with the sub-optimal exponential tilting measure in (26) for the (8,16) regular LDPC code ensembles of Gallager with block lengths of 1008 and 10080 symbols. The bounds are plotted in Figure 7 jointly with union bounds 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 decoding error probabilities 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 about 2.0 dB for a block error rate of 10^{-4}. For a block length of 10080 symbols, this gap is reduced to about 1.5 dB. Note that the use of the upper bound in Theorem 4 closes the 3 dB gap between the union upper bound and the respective ISP lower bound to only 1.5 dB while referring to a block length of 10080 symbols and a block error probability of 10^{-4}.

V. Summary and Conclusions

This paper considers the performance of non-binary linear block codes whose transmission takes place over memoryless symmetric channels. To this end, upper bounds on the decoding error probability are derived for finite-length codes. The general bounding approach is based on a partitioning of the original ensemble into two subsets of codebooks, according to their minimal Hamming distance: The performance of the set of codebooks with a relatively low minimum Hamming distance is assessed via a simple union bound which only depends on the considered ensemble, whereas the other set is evaluated using the second version of the Duman and Salehi (DS2) bound (See Section III-A). 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 to the non-binary setting is similar to the work of 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] where transmission takes place over the q-ary symmetric channel and the AWGN channel with a q-ary PSK modulation. The bound provided in Theorem 3 is attractive and show meaningful results up to the ultimate capacity limit. In addition, it outperforms the adaptation of the SFB in Theorem 2 for the non-binary setting which is even pronounced as the cardinality of the code alphabet is increased.

The weakness of the union bound is exemplified in this paper for regular LDPC code ensembles, showing the necessity in the replacement of the union bound with some improved upper bounds on the decoding error probability. On the other hand, 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 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 lower bound derived in [21]. These comparisons show by examples that recent sphere-packing bounds form a useful analytical tool for finite-length block codes.

Acknowledgment

The authors are grateful to Mark Flanagan for his insightful comments and his proof for Lemma 1. They are also grateful to Cong Ling for his feedback on an earlier version of this paper. We acknowledge the editor, Ludo Tolhuizen, for his observation in Remark 4 of Appendix F. The authors thank Gil Wiechman for some help with the calculation of the ISP bounds. The authors also acknowledge the anonymous reviewers and the associate editor for providing their feedback in a short time.

APPENDIX A
PROOF OF LEMMA 1

Let $x_1, x_2, x_3 \in \mathcal{X}$, p be the transition probability of the channel, and $\mathcal{T}$ be the mapping as in Lemma 1. Then, by setting $x \triangleq x_3 - x_2$, it follows from (1) that for all $y' \in \mathcal{Y}$

$$p(y'|x) = p\big(\mathcal{T}(y', x_2)|x_2 + x\big).$$

As a particular case, for $y' = \mathcal{T}(y, x_1)$ where $y \in \mathcal{Y}$, we have

$$p\big(\mathcal{T}(y, x_1)|x\big) = p\Big(\mathcal{T}\big(\mathcal{T}(y, x_1), x_2\big)|x_2 + x\Big). \tag{28}$$

Using (1) (repeatedly twice) on the LHS of (28) it follows that

$$p\big(\mathcal{T}(y, x_1)|x\big) = p(y|x - x_1) = p\big(\mathcal{T}(y, x_3 - x + x_1)|x_3\big). \tag{29}$$

which then yields from (28) and (29), jointly with the equality $x_3 - x = x_2$, that

$$p(\mathcal{T}(y, x_1 + x_2)|x_3) = p\Big(\mathcal{T}\big(\mathcal{T}(y, x_1), x_2\big)|x_3\Big)$$

which coincides with (2).

APPENDIX B
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

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

Using the change of variables

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

it follows that

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

where

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

Since the code $\mathcal{C}$ is a 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 for 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

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

which concludes the proof.

Appendix C
Proof of Proposition 2

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

$$\begin{aligned}
P_{\text{e}} &= P_{\text{e}|0} \leq \left(\sum_{\mathbf{y} \in \mathcal{Y}^n} G_n^0(\mathbf{y}) p_n(\mathbf{y}|\mathbf{0}) \right)^{1-\rho} \\
&\quad \cdot \left\{ \sum_{m' \neq 0} \sum_{\mathbf{y} \in \mathcal{Y}^n} G_n^0(\mathbf{y})^{1-\frac{1}{\rho}} p_n(\mathbf{y}|\mathbf{0}) \left(\frac{p_n(\mathbf{y}|\mathbf{x}_{m'})}{p_n(\mathbf{y}|\mathbf{0})} \right)^{\lambda} \right\}^{\rho}. \qquad (30)
\end{aligned}$$

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

$$P_{\text{e}} \leq \left(\sum_{\mathbf{y} \in \mathcal{Y}^n} \prod_{i=1}^{n} g(y_i) p(y_i|0) \right)^{1-\rho} \cdot \left\{ \sum_{m' \neq 0} \sum_{\mathbf{y} \in \mathcal{Y}^n} \prod_{i=1}^{n} g(y_i)^{1-\frac{1}{\rho}} p(y_i|0) \left(\frac{p(y_i|x_{m',i})}{p(y_i|0)} \right)^{\lambda} \right\}^{\rho} \qquad (31)$$

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

Appendix D
Proof of Theorem 2

From (8)

$$
\begin{aligned}
&\Pr(\text{ error } \mid d_{\min} > D_n)\\
&\le \Bigl(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\Bigr)^{n(1-\rho)} \qquad q^{-n\rho(1-R)}\\
&\quad\cdot\Biggl(\sum_{\mathbf{t}\in\mathcal{H}:\ n-t_0>D_n} \frac{\mathsf{E}\Bigl[\bigl|\mathcal{C}_{\mathbf{t}}\bigr| \Bigm| d_{\min} > D_n\Bigr]}{q^{-n(1-R)}\binom{n}{\mathbf{t}}}\binom{n}{\mathbf{t}}\prod_{x\in\mathcal{X}}\bigl(s_{\lambda,\rho}(x)\bigr)^{t_x}\Biggr)^{\rho}\\
&\le \Bigl(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\Bigr)^{n(1-\rho)} \qquad q^{-n\rho(1-R)}\\
&\quad\cdot\Biggl(\max_{\mathbf{t}\in\mathcal{H}:\ n-t_0>D_n}\Biggl\{\frac{\mathsf{E}\Bigl[\bigl|\mathcal{C}_{\mathbf{t}}\bigr| \Bigm| d_{\min} > D_n\Bigr]}{q^{-n(1-R)}\binom{n}{\mathbf{t}}}\Biggr\}\Biggr)^{\rho}\\
&\quad\cdot\Biggl(\sum_{\mathbf{t}\in\mathcal{H}:\ n-t_0>D_n}\binom{n}{\mathbf{t}}\prod_{x\in\mathcal{X}}\bigl(s_{\lambda,\rho}(x)\bigr)^{t_x}\Biggr)^{\rho}
\end{aligned}
$$

where the last transition holds since $\sum_i x_i y_i \le \max_i x_i \sum_i y_i$ if $\{x_i\}$ and $\{y_i\}$ are non-negative sequences. Let $\mathcal{X}^* \triangleq \mathcal{X}\setminus\{0\}$, from the definition of α_q in (13) we get

$$
\begin{aligned}
&\Pr(\text{ error } \mid d_{\min} > D_n)\\
&\le q^{-n\rho(1-R)}\ \Bigl(\alpha_q(\mathcal{C}, D_n)\Bigr)^{\rho}\ \Bigl(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\Bigr)^{n(1-\rho)}\\
&\quad\cdot\Biggl[\sum_{l=D_n+1}^{n}\binom{n}{l}\bigl(s_{\lambda,\rho}(0)\bigr)^{n-l}\sum_{t_1+\ldots+t_{q-1}=l}\binom{l}{t_1,\ldots,t_{q-1}}\prod_{x\in\mathcal{X}^*}\bigl(s_{\lambda,\rho}(x)\bigr)^{t_x}\Biggr]^{\rho}\\
&= q^{-n\rho(1-R)}\ \Bigl(\alpha_q(\mathcal{C}, D_n)\Bigr)^{\rho}\ \Bigl(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\Bigr)^{n(1-\rho)}\\
&\quad\cdot\Biggl[\sum_{l=D_n+1}^{n}\binom{n}{l}\bigl(s_{\lambda,\rho}(0)\bigr)^{n-l}\Bigl(\sum_{x\in\mathcal{X}^*}s_{\lambda,\rho}(x)\Bigr)^{l}\Biggr]^{\rho}\\
&\le q^{-n\rho(1-R)}\ \Bigl(\alpha_q(\mathcal{C}, D_n)\Bigr)^{\rho}\ \Bigl(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\Bigr)^{n(1-\rho)}\\
&\quad\cdot\Biggl[\sum_{l=0}^{n}\binom{n}{l}\bigl(s_{\lambda,\rho}(0)\bigr)^{n-l}\Bigl(\sum_{x\in\mathcal{X}^*}s_{\lambda,\rho}(x)\Bigr)^{l}\Biggr]^{\rho}\\
&\le q^{-n\rho(1-R)}\ \Bigl(\alpha_q(\mathcal{C}, D_n)\Bigr)^{\rho}\ \Bigl(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\Bigr)^{n(1-\rho)}\ \Bigl(\sum_{x\in\mathcal{X}}s_{\lambda,\rho}(x)\Bigr)^{n\rho}.
\end{aligned}
\tag{32}
$$

Next, setting

$$
g(y) = \Bigl(\frac{1}{q}\sum_{x\in\mathcal{X}}p(y|x)^{\frac{1}{1+\rho}}\Bigr)^{\rho}\ p(y|0)^{-\frac{\rho}{1+\rho}},\quad \lambda = \frac{1}{1+\rho} \tag{33}
$$

it follows that

$$\sum_{y\in\mathcal{Y}} g(y)p(y|0) = \sum_{y\in\mathcal{Y}} \left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{\rho} p(y|0)^{\frac{1}{1+\rho}}. \tag{34}$$

In addition, plugging (33) in (9), we get

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

which then implies from (34) that

$$\begin{aligned}\sum_{x\in\mathcal{X}} s_{\lambda,\rho}(x) &= q\sum_{y\in\mathcal{Y}} \left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{\rho} p(y|0)^{\frac{1}{1+\rho}}\\ &= q\sum_{y\in\mathcal{Y}} g(y)p(y|0).\end{aligned} \tag{35}$$

From (32) and (35), it follows that

$$\Pr(\text{ error } \mid d_{\min} > D_n) \le q^{n\rho R}\,\Big(\alpha_q(\mathcal{C},D_n)\Big)^{\rho}\,\left(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\right)^{n}. \tag{36}$$

To complete the proof, we need the following lemma:

Lemma 6. Setting g(y) as in (33), the following equality follows for all ξ:

$$\sum_{y\in\mathcal{Y}} g(y)^{\xi}p(y|0) = \sum_{y\in\mathcal{Y}} \left[\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{\xi\rho}\cdot\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{1-\frac{\xi\rho}{1+\rho}}\right)\right]. \tag{37}$$

Proof: Since the channel 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 (33) we have

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

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

$$p(\mathcal{T}(\mathcal{T}(y,x'),-x')|x) = p(y|x) \tag{38}$$

for all $x, x' \in \mathcal{X}$ and $y \in \mathcal{Y}$. Next, using the closure of the (finite) input alphabet, it follows that

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

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. ■

From (36) and Lemma 6 (with $\xi = 1$ in (37)), we get after an optimization over ρ (where $0 \le \rho \le 1$):

$$\Pr(\text{ error } \mid d_{\min} > D_n) \le q^{-nE_{\mathrm{r}}\left(R+\frac{\log_q \alpha_q(\mathcal{C},D_n)}{n}\right)}. \tag{39}$$

Finally, the proof of Theorem 2 follows from Lemma 3 and (39).

Appendix E
Proof of Theorem 3

Under the conditions in Theorem 3, we get from (8), (14), and (15) that

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

where $\mathcal{X}^* \triangleq \mathcal{X}\setminus\{0\}$. Next, setting λ and $g(y)$ as defined in (33), then it follows from (35) that

$$\begin{aligned}
\Pr(\text{ error } \mid d_{\min} > D_n) \le{}& \left(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\right)^{n(1-\rho)}\\
&\cdot\left[\sum_{n-t_0>D_n}\frac{P(n-t_0)}{1-\epsilon_n}\binom{n}{t_0}\big(s_{\lambda,\rho}(0)\big)^{t_0}\left(q\sum_{y\in\mathcal{Y}} g(y)p(y|0)-s_{\lambda,\rho}(0)\right)^{n-t_0}\right]^{\rho}.
\end{aligned}\tag{40}$$

The proof is completed by applying Lemma 6 in (40) with $\xi = 1$ for $\sum_{y\in\mathcal{Y}} g(y)p(y|0)$, and with $\xi = 1-\frac{1}{\rho}$ for $s_{\lambda,\rho}(0) = \sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}}p(y|0)$.

APPENDIX F
PROOF OF LEMMA 5

Denote by $a_{x^*}(l)$ the number of choices of l non-zero elements in $\{1,\ldots,q-1\}$ whose summation modulo q equals x^* (where $x^* \in \{0,\ldots,q-1\}$). Note that these l elements should not necessarily be distinct. Then, for $1 \le l \le d$, there are $\binom{d}{l} a_{x^*}(l)$ vectors $\mathbf{x} = (x_1,\ldots,x_d)$, whose Hamming weight is l, and which satisfy

$$x_1 + \cdots + x_d = x^* \bmod q. \tag{41}$$

The sequences $\{a_{x^*}(l)\}$ satisfy the following system of recursive equations:

$$a_{x^*}(l) = \sum_{x=1}^{q-1} a_{(x^*-x) \bmod q}(l-1), \quad x^* = 0,1,\ldots,q-1 \tag{42}$$

with the initial conditions $a_0\,(1) = 0$, and $a_x\,(1) = 1$ for $x \in \{1,\ldots,q-1\}$. Using a vector notation, the equations in (42) are written as

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

whose solution for $l \ge 1$ is given by

$$\begin{pmatrix} a_0(l) \\ a_1(l) \\ \vdots \\ \\ a_{q-1}(l) \end{pmatrix} = \begin{pmatrix} 0 & 1 & \cdots & 1 & 1 \\ 1 & 0 & 1 & \cdots & 1 \\ \vdots & & \ddots & & \\ 1 & \cdots & 1 & 0 & 1 \\ 1 & & \cdots & 1 & 0 \end{pmatrix}_{q\times q}^{l-1} \begin{pmatrix} 0 \\ 1 \\ \vdots \\ \\ 1 \end{pmatrix}_{q\times 1} . \tag{43}$$

In proving the considered lemma, the main ingredient is obtaining the number of vectors $\mathbf{x}$ satisfying the parity-check equation

$$x_1 + \cdots + x_d = 0 \bmod q. \tag{44}$$

Accordingly, only the sequence $\{a_0(l)\}$ is of interest. To obtain a closed form expression for this sequence, consider the following difference equation:

$$\begin{cases} u_l = (q-1)\big(u_{l-1} + (-1)^l\big) \\ u_1 = 0 \end{cases} . \tag{45}$$

It can be verified by induction that the elements on the diagonal of the $q \times q$ matrix on the RHS of (43), raised to the $(l-1)$-th power, are identical and equal to u_{l-1}, where the sequence $\{u_l\}$ is the solution of (45). Moreover, all other elements outside the diagonal, are equal to $u_{l-1} + (-1)^l$. As a result, it follows from (43) that

$$a_0(l) = (q-1)\left(u_{l-1} + (-1)^l\right),\ l \ge 1,\ a_0(1) = 0.$$

which implies from (45) that $a_0(l) = u_l$ for $l \ge 1$. Solving the difference equation in (45), gives

$$a_0(l) = \frac{(q-1)^l + (q-1)(-1)^l}{q}, \quad l \ge 1.$$

Hence, the enumerator for the number of vectors $\mathbf{x}$ satisfying the parity-check equation in (44), is given by $A^*(X)$ in (21). As a result, the enumerator of the first sub-matrix in the considered ensemble is given in (20) (this is similar to the idea provided in [31] for the binary case). Finally, (19) is established in [22] which concludes the proof of Lemma 5.

Remark 4. The weight enumerator of the single parity-check (SPC) code (as specified in (41)) can be alternatively derived via Mac Williams' Theorem for non-binary linear block codes (see [41, Theorem 4.6]). Since the dual of a

SPC code is a repetition code, then the above result follows easily. Note however that the Mac Williams' theorem applies to the case where q is an integral power of a prime number (which forms a necessary and sufficient condition for the existence of a Galois field of size q) which is not required in this lemma. Since a repetition code is a maximum distance separable (MDS) code then various properties hold in fact over alphabets of arbitrary size (see [42]).

APPENDIX G
PROOF OF THEOREM 4

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

$$
\begin{aligned}
&P(\text{ error } | d_{\min} > D_n) \\
&= \mathsf{E}\left[\iint_{(\mathbf{y},\mathbf{a}): p(\mathbf{y},\mathbf{a}|\mathbf{x}) \geq p(\mathbf{y},\mathbf{a}|\mathbf{0}) \text{ for some } \mathbf{x}\neq\mathbf{0}\in\mathcal{C}} p(\mathbf{y},\mathbf{a}|\mathbf{0}) \; d\mathbf{y}\, d\mathbf{a} \,\Big|\, d_{\min} > D_n \right] \\
&\leq \mathsf{E}\left[\iint_{\mathbf{y},\mathbf{a}} p(\mathbf{y},\mathbf{a}|\mathbf{0}) \sum_{j=1}^{J} \left(\sum_{\mathbf{t}\in\mathcal{H}_j} \sum_{\mathbf{x}\in\mathcal{C}_{\mathbf{t}}} \left(\frac{p(\mathbf{y},\mathbf{a}|\mathbf{x})}{p(\mathbf{y},\mathbf{a}|\mathbf{0})} \right)^{\lambda_j} \right)^{\rho_j} d\mathbf{y}\, d\mathbf{a} \,\Big|\, d_{\min} > D_n \right] \\
&= \sum_{j=1}^{J} \mathsf{E}\Bigg[\iint_{\mathbf{y},\mathbf{a}} \psi_j(\mathbf{y},\mathbf{a}) \\
&\qquad \left(\sum_{\mathbf{t}\in\mathcal{H}_j} \sum_{\mathbf{x}\in\mathcal{C}_{\mathbf{t}}} \psi_j(\mathbf{y},\mathbf{a})^{-\frac{1}{\rho_j}} p(\mathbf{y},\mathbf{a}|\mathbf{0})^{\frac{1-\lambda_j\rho_j}{\rho_j}} p(\mathbf{y},\mathbf{a}|\mathbf{x})^{\lambda_j} \right)^{\rho_j} d\mathbf{y}\, d\mathbf{a} \,\Big|\, d_{\min} > D_n \Bigg] \qquad (46)
\end{aligned}
$$

where the statistical expectation is taken over all the codebooks whose Hamming minimum distance is larger than D_n. From (46), using Jensen's inequality we have

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

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

$$
\begin{aligned}
&P(\text{ error } | d_{\min} > D_n) \\
&\leq \sum_{j=1}^{J} \mathsf{E}\left[\left(\sum_{\mathbf{t}\in\mathcal{H}_j} \sum_{\mathbf{x}\in\mathcal{C}_{\mathbf{t}}} \iint_{\mathbf{y},\mathbf{a}} \prod_{i=1}^{n} \psi_j(y_i,a_i)^{-\frac{1}{\rho_j}} p(y_i,a_i|0)^{\frac{1-\lambda_j\rho_j}{\rho_j}} p(y_i,a_i|x_i)^{\lambda_j}\, dy_i\, da_i \right)^{\rho_j} \,\Big|\, d_{\min} > D_n \right] \\
&= \sum_{j=1}^{J} \mathsf{E}\left[\left(\sum_{\mathbf{t}\in\mathcal{H}_j} |\mathcal{C}_{\mathbf{t}}| \prod_{x\in\mathcal{X}} \left(\iint_{y,a} \psi_j(y,a)^{1-\frac{1}{\rho_j}} p(y,a|0)^{\frac{1-\lambda_j\rho_j}{\rho_j}} p(y,a|x)^{\lambda_j}\, dy\, da \right)^{t_x} \right)^{\rho_j} \,\Big|\, d_{\min} > D_n \right].
\end{aligned}
$$

The proof is concluded by using Jensen's inequality (for the statistical expectation) and Lemma 3.

APPENDIX H
A CLOSED-FORM EXPRESSION FOR THE INTEGRAL IN THEOREM 4 WHEN APPLIED TO EXAMPLE 7

Similarly to [9], we will pursue a closed-form expression by examining an exponential tilting probability measure ψ as in (26). Note that the joint pdf $p(y,a|x)$ 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

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

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 $\mu(x) \triangleq \sqrt{\frac{2E_{\text{s}}}{N_0}} \exp\left(\frac{2\pi i}{q}\, x\right)$ is the q-ary PSK modulation mapping applied in the considered scheme. In addition, ψ_j in (26) is easily verified to be a probability measure. Assuming that $1+K+\beta > 0$ (which is the case since $\alpha \geq 0$), the denominator of ψ as in (26) equals

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

Straightforward (though tedious) calculations show that for every $x \in \mathcal{X}$

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

where

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

APPENDIX I
A CLOSED-FORM EXPRESSION FOR THE INTEGRAL IN THEOREM 4 WHEN APPLIED TO EXAMPLE 8

The following exponential tilting measure is applied:

$$\psi(y,a) = \frac{\alpha p(a)}{2\pi} \left(1 + \frac{\beta}{L}\sqrt{\frac{2E_s}{N_0}}\right)^L \exp\left(-\frac{\alpha}{2}\left|y - a\sqrt{\frac{2E_{\text{s}}}{N_0}}u\right|^2 - \beta a^2 \sqrt{\frac{2E_{\text{s}}}{N_0}}\right) \tag{47}$$

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 (27). The integral in (24) with the proposed tilting measure in (47) is calculated via straightforward calculus, and it is obtained that for every $x \in \mathcal{X}$

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

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, February 2007.

[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 belief 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] 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.

[27] 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.

[28] T. Duman and M. Salehi, “Performance bounds for turbo-coded modulation systems,” *IEEE Trans. Communications*, vol. 47, no. 4, pp. 511-521, April 1999.

[29] 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.

[30] 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.

[31] S. Tong, “Tangential-sphere bounds on the ensemble performance of ML decoded Gallager codes via their exact ensemble distance spectrum,” *Proceeding of the 2008 IEEE International Conference on Communications (ICC 2008)*, pp. 1150–1154, Beijing, China, May 2008.

[32] D. Burshtein and G. Miller, “Asymptotic enumeration methods for analyzing LDPC codes,” *IEEE Trans. on Information Theory*, vol. 50, no. 6, pp. 1115–1131, June 2004.

[33] A. J. Viterbi and J. K. Omura, *Principle of Digital Communication and Coding*, 1979.

[34] M. F. Flanagan, V. Skachek, E. Byrne and M. Greferath, “Linear-programming decoding of non-binary linear codes,” *Proceeding of the 7th International ITG Conference on Source and Channel Coding,* Ulm, Germany, January 2008.

[35] M. Flanagan, “Codeword-independent performance of nonbinary linear codes under linear-programming and sum-product decoding,” *Proceedings 2008 IEEE International Symposium on Information Theory (ISIT ’08)*, pp. 1503–1507, Toronto, Canada, July 2008.

[36] 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.

[37] 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.

[38] 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.

[39] 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.

[40] D. E. Knuth, *The Art of Computer Programming.* Vol.2: Seminumerical Algorithms, (3rd edition), pp. 461, Addison-Wesley, 1998.

[41] R. M. Roth, *Introduction to Coding Theory*, Cambridge University Press, 2006.

[42] L. M. G. M. Tolhuizen, “On maximum distance separable codes over alphabets of arbitrary size,” *Proceedings 1994 IEEE International Symposium on Information Theory (ISIT ’94)*, p. 431, Trondheim, Norway, June 1994.