Improvement of Gallager Upper Bound and its Variations for Discrete Channels

Kostis Xenoulis, *Student Member, IEEE*, and Nicholas Kalouptsidis, *Senior Member, IEEE*

***Abstract*—A new tight upper bound on the maximum-likelihood (ML) word and bit-error decoding probabilities for specific codes over discrete channels is presented. It constitutes an enhanced version of the Gallager upper bound and its variations resulting from the Duman–Salehi second bounding technique. An efficient technique is developed that, in the case of symmetric channels, overcomes the difficulties associated with the direct computation of the proposed bound. Surprisingly, apart from the distance and input–output weight enumerating functions (IOWEFs), the bound depends also on the coset weight distribution of the code.**

***Index Terms*—Coset weight and distance distributions, Gallager bound, inverse exponential sum inequality, maximum-likelihood (ML) decoding, word and bit error probabilities.**

I. Introduction

ERROR probability evaluation is a significant performance measure of coded information transmission over various communication channels. The high complexities involved in its calculation necessitates the introduction of efficient bounding techniques. Classical treatments [1] as well as modern approaches [2, pp. 2–3] provide tight bounds mostly for random and specific families of codes (turbo codes [3], low-density parity-check (LDPC) codes [4]), since the latter are treated more easily than specific codes. Thus, the existence of at least one optimum code within these families is assured, but the respective characteristics of the optimum code remain unknown. The development of new bounding techniques is crucial to the accommodation of optimum specific codes, which can achieve arbitrarily low error decoding probability with rates close to the channel's capacity.

The present work introduces a tight upper bound on the maximum-likelihood (ML) error decoding probability of specific codes over discrete channels. The bound is deduced from an inverse sum exponential inequality, analogous to the log-sum inequality [5, Theorem (2.7.1)], over the set of erroneous received vectors and is suitable for both word and bit error probability analysis. Moreover, the bound is proved to be a tighter version of Gallager's first upper bound [1]. Due to the above characteristics, application of the DS2 bounding technique leads to a tighter version of the DS2 bound [6]. As a consequence, the proposed bound produces tighter bounds than several known cases of the DS2 form, such as Divsalar [7].

Manuscript received June 21, 2008; revised May 18, 2009. Current version published August 19, 2009.

The authors are with the Department of Informatics and Telecommunications, University of Athens, Athens 15784, Greece (e-mail: kxen@di.uoa.gr; kalou@di.uoa.gr).

Communicated by G. Seroussi, Associate Editor for Coding Theory.
Digital Object Identifier 10.1109/TIT.2009.2025576

The direct calculation of the proposed bound is prohibitively complex for specific large length codes. For instance, if linear block codes are used for information transmission through a binary-symmetric channel, the determination of the corresponding Voronoi regions [8] is necessary for the bound calculation. As a consequence, the majority of performance bounds existing in the literature [2, pp. 3–4] rely only on basic features, such as the distance spectrum and input–ouput weight enumeration function (IOWEF) of the examined codes.

In the rest of this paper, Section II introduces the tight upper bound on the ML word and bit error decoding probabilities, respectively, based on the inverse exponential sum inequality. Section III treats the set of erroneous received vectors for discrete symmetric channels and thus closed-form expressions of the bound are obtained. Section IV concludes the analysis and applies the new upper bound to specific codes, thus revealing its tightness over the DS2 bound.

II. Tight Upper Bound on the ML Error Decoding Probability

Let C be a block code of length n and dimension k, over a field F with q elements. Let also $\boldsymbol{c}_i \in F^n, i = 0, \ldots, q^k - 1$ and $S_d, d = d_{\min}, \ldots, n$ denote, respectively, the codewords and the distance distribution of the code C, with $d_{\min}$ its minimum distance. If $\boldsymbol{c}$ and $\boldsymbol{y}$ are two vectors in F^n, then $d_H(\boldsymbol{c}, \boldsymbol{y})$ is the Hamming distance between them and $\mathrm{wt}(\boldsymbol{c}), \mathrm{wt}(\boldsymbol{y})$ their corresponding Hamming weights. For an arbitrary set of messages $\mathcal{M}$ with cardinality $|\mathcal{M}|$, a message $m, 0 \leq m \leq |\mathcal{M}| - 1$, is mapped to a codeword $\boldsymbol{c}_m$ of the above code C and is transmitted over a discrete communication channel with transition probability $\Pr(\boldsymbol{y}|\boldsymbol{c}_m)$. $\boldsymbol{y}$ is the received vector at the output of the channel, also of length n. The set of received vectors is denoted by $\boldsymbol{Y}$. Each received vector $\boldsymbol{Y}$ is decoded back onto the set of messages $\mathcal{M}$, according to the ML rule. For the aforementioned transmission procedure, Gallager's upper bound [1] on the code's error decoding probability yields

$$P_{e|m} \leq \sum_{\boldsymbol{y} \in \boldsymbol{Y}} \Pr(\boldsymbol{y}|\boldsymbol{c}_m) \mathcal{K}_m(\boldsymbol{y}, C, \lambda, \rho) \tag{1}$$

where

$$\mathcal{K}_m(\boldsymbol{y}, C, \lambda, \rho) = \left(\sum_{m' \neq m} \left(\frac{\Pr(\boldsymbol{y}|\boldsymbol{c}_{m'})}{\Pr(\boldsymbol{y}|\boldsymbol{c}_m)} \right)^{\lambda} \right)^{\rho} \tag{2}$$

and $\lambda, \rho \geq 0$. A modified version is provided by the DS2 technique [2, Sec. 4.2.2]. Let $G_m(\boldsymbol{y})$ be an arbitrary nonnegative function over $\boldsymbol{Y}$, that may also depend on the transmitted message m. Then, for $\lambda \geq 0$ and $0 \leq \rho \leq 1$

$$P_{e|m} \leq \left(\sum_{\boldsymbol{y}\in\boldsymbol{Y}} \Pr(\boldsymbol{y}|\boldsymbol{c}_m) G_m(\boldsymbol{y})\right)^{1-\rho} \cdot \left(\sum_{m'\neq m}\sum_{\boldsymbol{y}\in\boldsymbol{Y}} \Pr(\boldsymbol{y}|\boldsymbol{c}_m) G_m(\boldsymbol{y})^{1-1/\rho} \left(\frac{\Pr(\boldsymbol{y}|\boldsymbol{c}_{m'})}{\Pr(\boldsymbol{y}|\boldsymbol{c}_m)}\right)^{\lambda}\right)^{\rho}. \tag{3}$$

The introduction of a tighter upper bound on the ML error decoding probability is made possible by the following inverse exponential sum inequality.

Theorem 1 (Inverse Exponential Sum Inequality): For positive numbers, $\alpha_1, \alpha_2, \ldots, \alpha_n$ and $\beta_1, \beta_2, \ldots, \beta_n$

$$\sum_{i=1}^{n} \beta_i \leq \left(\sum_{i=1}^{n} \alpha_i\right) \ln \frac{\sum_{i=1}^{n} \alpha_i e^{\beta_i/\alpha_i}}{\sum_{i=1}^{n} \alpha_i} \tag{4}$$

with equality if and only if $\beta_i/\alpha_i = \text{const}$.

Proof: The function $f(t) = te^{t^{-1}}, t > 0$ is strictly convex on the set of positive reals because $f''(t) = t^{-3}e^{t^{-1}} > 0$ for all positive t. Hence, by Jensen's inequality, it holds that

$$\sum_i a_i t_i e^{t_i^{-1}} \geq \left(\sum_i a_i t_i\right) e^{\left(\sum_i a_i t_i\right)^{-1}} \tag{5}$$

with $a_i \geq 0$ and $\sum_i a_i = 1$. Setting $t_i = \alpha_i/\beta_i$ and $a_i = \beta_i/\sum_i \beta_i$ in (5), the later is transformed into

$$\sum_i \frac{\beta_i}{\left(\sum_i \beta_i\right)} \frac{\alpha_i}{\beta_i} e^{\beta_i/\alpha_i} \geq \left(\sum_i \frac{\beta_i}{\left(\sum_i \beta_i\right)} \frac{\alpha_i}{\beta_i}\right) \cdot e^{\left(\sum_i \beta_i/\left(\sum_i \beta_i\right)\alpha_i/\beta_i\right)^{-1}} \tag{6}$$

or, equivalently, into

$$\sum_i \alpha_i e^{\beta_i/\alpha_i} \geq \left(\sum_i \alpha_i\right) e^{\sum_i \beta_i / \sum_i \alpha_i}. \tag{7}$$

Rearranging and taking logarithms in (7), we establish (4). □

The inverse exponential sum inequality of Theorem 1 is used below in the error decoding probability analysis.

Theorem 2: Consider the transmission of an arbitrary set of messages $\mathcal{M}$ over a discrete communication channel, through the utilization of an (n, k) code C. Let $\boldsymbol{Y}_m^b$ denote the set of erroneous received vectors given that the message m is transmitted

$$\boldsymbol{Y}_m^b = \Big\{\boldsymbol{y} \in \boldsymbol{Y} : \exists m' \in \mathcal{M},\ m' \neq m, \Pr(\boldsymbol{y}|\boldsymbol{c}_{m'}) \geq \Pr(\boldsymbol{y}|\boldsymbol{c}_m)\Big\}$$

and

$$\mathcal{L}_m(C, \lambda, \rho) = \min_{\boldsymbol{y}\in\boldsymbol{Y}_m^b} \mathcal{K}_m(\boldsymbol{y}, C, \lambda, \rho).$$

Then the ML word error decoding probability for the specific code, given that the message m is transmitted, is upper-bounded for all $\lambda, \rho \geq 0$ by

$$P_{e|m} \leq \mathcal{L}_m(C, \lambda, \rho)^{-1} \left(\sum_{\boldsymbol{y}\in\boldsymbol{Y}} \Pr(\boldsymbol{y}|\boldsymbol{c}_m) \mathcal{K}_m(\boldsymbol{y}, C, \lambda, \rho)\right) \leq \sum_{\boldsymbol{y}\in\boldsymbol{Y}} \Pr(\boldsymbol{y}|\boldsymbol{c}_m) \mathcal{K}_m(\boldsymbol{y}, C, \lambda, \rho). \tag{8}$$

Proof: The leftmost inequality is derived by invoking Theorem 1. Indeed, for each $\boldsymbol{y} \in \boldsymbol{Y}_m^b$, set

$$\beta_{\boldsymbol{y}} = \Pr(\boldsymbol{y}|\boldsymbol{c}_m),\ \alpha_{\boldsymbol{y}} = \beta_{\boldsymbol{y}} \mathcal{K}_m(\boldsymbol{y}, C, \lambda, \rho). \tag{9}$$

Then

$$\sum_{\boldsymbol{y}\in\boldsymbol{Y}_m^b} \Pr(\boldsymbol{y}|\boldsymbol{c}_m) = P_{e|m}. \tag{10}$$

Application of Theorem 1 leads to

$$P_{e|m} \leq \ln \frac{\sum_{\boldsymbol{y}\in\boldsymbol{Y}_m^b} \Pr(\boldsymbol{y}|\boldsymbol{c}_m) \mathcal{K}_m(\boldsymbol{y}, C, \lambda, \rho) e^{\mathcal{K}_m(\boldsymbol{y}, C, \lambda, \rho)^{-1}}}{\sum_{\boldsymbol{y}\in\boldsymbol{Y}_m^b} \Pr(\boldsymbol{y}|\boldsymbol{c}_m) \mathcal{K}_m(\boldsymbol{y}, C, \lambda, \rho)} \times \left(\sum_{\boldsymbol{y}\in\boldsymbol{Y}_m^b} \Pr(\boldsymbol{y}|\boldsymbol{c}_m) \mathcal{K}_m(\boldsymbol{y}, C, \lambda, \rho)\right). \tag{11}$$

The claim follows if in the first term $\mathcal{K}_m(\boldsymbol{y}, C, \lambda, \rho)$ is replaced by its minimum value $\mathcal{L}_m(C, \lambda, \rho)$ over $\boldsymbol{Y}_m^b$ and the sum of the second term in (11) is extended over the larger set $\boldsymbol{Y}$. To establish the second inequality, it suffices to show that the term involving the logarithm is less than 1. This is indeed the case because $\mathcal{L}_m(C, \lambda, \rho)^{-1}$ is lower or equal to 1. □

Theorem 2 provides a bound on the ML word error decoding probability that is tighter than Gallager bound, as noted from the second inequality in (8). Moreover, the DS2 technique can be applied to the second term of the first inequality in (8) for all $\rho \leq 1$, thus leading to a tighter version of the DS2 bound.

Theorem 3: Under the assumptions of Theorem 2, the ML word error decoding probability is upper-bounded for all $\lambda \geq 0, 0 \leq \rho \leq 1$ and any nonnegative function $G_m(\boldsymbol{y})$ by

$$P_{e|m} \leq \mathcal{L}_m(C, \lambda, \rho)^{-1} \left(\sum_{\boldsymbol{y}\in\boldsymbol{Y}} \Pr(\boldsymbol{y}|\boldsymbol{c}_m) G_m(\boldsymbol{y})\right)^{1-\rho} \cdot \left(\sum_{m'\neq m}\sum_{\boldsymbol{y}\in\boldsymbol{Y}} \Pr(\boldsymbol{y}|\boldsymbol{c}_m) G_m(\boldsymbol{y})^{1-1/\rho} \left(\frac{\Pr(\boldsymbol{y}|\boldsymbol{c}_{m'})}{\Pr(\boldsymbol{y}|\boldsymbol{c}_m)}\right)^{\lambda}\right)^{\rho}. \tag{12}$$

A similar upper bound on the bit error decoding probability can easily be deduced as a special case of theorem 1. Specifically, as in [7], consider a binary systematic (n, k) block code C, and an encoder that maps each k-bit information block into an n-bit codeword $\boldsymbol{c} \in C$. For each codeword $\boldsymbol{c}_m \in C$, the corresponding information block $\boldsymbol{u}_m$ has components $u_{m,i}, i = 1, \ldots, k$ respectively. Then, the bit error decoding probability given the transmitted message m, is written as

$$P_{b|m} = \frac{1}{k} \sum_{\boldsymbol{y}\in\boldsymbol{Y}_m^{b\prime}} \Pr(\boldsymbol{y}|\boldsymbol{c}_m) \tag{13}$$

where

$$\boldsymbol{Y}_m^{b\,\prime} = \Big\{\boldsymbol{y} \in \boldsymbol{Y} : \exists\, m' \in \mathcal{M}, m' \neq m \text{ and } \exists\, i \in [1,k] \text{ such that } \Pr(\boldsymbol{y}|\boldsymbol{c}_{m'}) \geq \Pr(\boldsymbol{y}|\boldsymbol{c}_m) \text{ and } \boldsymbol{u}_{m',i} \neq \boldsymbol{u}_{m,i}\Big\}. \quad (14)$$

According to [7, eqs. (75)–(81)], when $\boldsymbol{y} \in \boldsymbol{Y}_m^{b'}$, it holds for all $0 \leq \rho \leq 1$ and $\lambda \geq 0$

$$\mathcal{K}_{m'}(\boldsymbol{y}, C, U, \lambda, \rho) \triangleq \left(\sum_{m' \neq m} d(\boldsymbol{u}_{m'}, \boldsymbol{u}_m)\left(\frac{\Pr(\boldsymbol{y}|\boldsymbol{c}_{m'})}{\Pr(\boldsymbol{y}|\boldsymbol{c}_m)}\right)^{\lambda}\right)^{\rho} \geq 1 \quad (15)$$

where U is the set of information blocks. Consequently, if β_i and α_i in Theorem 1 are replaced by $\Pr(\boldsymbol{y}|\boldsymbol{c}_m)/k$ and $\mathcal{K}'_m(\boldsymbol{y}, C, U, \lambda, \rho)\Pr(\boldsymbol{y}|\boldsymbol{c}_m)/k$, respectively, the following upper bound is obtained for the bit error decoding probability.

Theorem 4: If the assumptions of Theorem 2 hold and

$$\mathcal{L}'_m(C, U, \lambda, \rho) = \min_{\boldsymbol{y} \in \boldsymbol{Y}_m^{b\,\prime}} \mathcal{K}'_m(\boldsymbol{y}, C, U, \lambda, \rho) \quad (16)$$

then the ML bit error decoding probability satisfies

$$P_{b|m} \leq \mathcal{L}'_m(C, U, \lambda, \rho)^{-1} \cdot \left(\sum_{\boldsymbol{y} \in \boldsymbol{Y}} \frac{1}{k}\Pr(\boldsymbol{y}|\boldsymbol{c}_m)\mathcal{K}'_m(\boldsymbol{y}, C, U, \lambda, \rho)\right). \quad (17)$$

The second term of the product in the right-hand side of (17) equals Gallager's upper bound on the bit error probability while the first term of the same product is lower or equal to 1 due to (15).

III. Special Cases of Theorems 2 and 4

The transmission of an arbitrary set of messages $\mathcal{M}$ through a linear block code C, based on a field F with q elements, over memoryless and output-symmetric channels is considered in this section. Suppose that $(1-p)$ is the probability that a transmitted symbol is received correctly while $p/(q-1)$ is the probability that a transmitted symbol is converted into one of the $q-1$ remaining symbols. It is assumed that $1-p > p/(q-1)$. Without loss of generality, the all-zero codeword $\boldsymbol{c}_0$ is transmitted, and ML decoding is performed at the channel output. Due to the above transmission characteristics, ML decoding is equivalent to minimum distance decoding and the distance distribution $\{S_d\}$ of C equals its weight distribution. $\boldsymbol{Y}_0^b$ denotes the set of erroneous received vectors, given that $\boldsymbol{c}_0$ is transmitted.

A. Binary Input–Discrete Output Channels

Let $\mathcal{I} = \{0,1\}$ be the channels's input alphabet and $\mathcal{J}$ the channel's output alphabet, not necessarily binary. Let $\boldsymbol{y} \in \boldsymbol{Y}_0^b$. Then $\boldsymbol{y}$ is closer according to the ML decoding rule to a codeword $\boldsymbol{c}_m, m \neq 0$ than any other codeword, so that

$$\mathcal{K}_0(\boldsymbol{y}, C, \lambda, \rho) = \left[\left(\frac{\Pr(\boldsymbol{y}|\boldsymbol{c}_m)}{\Pr(\boldsymbol{y}|\boldsymbol{c}_0)}\right)^{\lambda} + \sum_{m' \neq m, 0}\left(\frac{\Pr(\boldsymbol{y}|\boldsymbol{c}_{m'})}{\Pr(\boldsymbol{y}|\boldsymbol{c}_0)}\right)^{\lambda}\right]^{\rho} \geq \left[1 + \sum_{m' \neq m, 0}\left(\frac{\Pr(\boldsymbol{y}|\boldsymbol{c}_{m'})}{\Pr(\boldsymbol{y}|\boldsymbol{c}_0)}\right)^{\lambda}\right]^{\rho}. \quad (18)$$

Since the channel is memoryless

$$\frac{\Pr(\boldsymbol{y}|\boldsymbol{c}_{m'})}{\Pr(\boldsymbol{y}|\boldsymbol{c}_0)} = \prod_{l=1}^{\mathrm{wt}(\boldsymbol{c}_{m'})}\frac{\Pr(y_{i_l}|1)}{\Pr(y_{i_l}|0)} \quad (19)$$

where $i_l, l = 1 \ldots \mathrm{wt}(\boldsymbol{c}_{m'})$ denote the positions of 1's in the codeword $\boldsymbol{c}_{m'}$. Equation (19) is always lower-bounded by

$$\left(\min_{y \in \mathcal{J}}\frac{\Pr(y|1)}{\Pr(y|0)}\right)^{\mathrm{wt}(\boldsymbol{c}_{m'})} \quad (20)$$

since each fraction in the right-hand side of (18) is lower-bounded by $\mathcal{X} = \min_{y \in \mathcal{J}} \Pr(y|1)/\Pr(y|0)$ and the number of fractions appearing in the corresponding expression is exactly $\mathrm{wt}(\boldsymbol{c}_{m'})$. Applying (20) to every term of the sum in the right-hand side of (18), we obtain

$$\mathcal{K}_0(\boldsymbol{y}, C, \lambda, \rho) \geq \left[1 + \sum_{m' \neq m, 0}\mathcal{X}^{\lambda \mathrm{wt}(\boldsymbol{c}_{m'})}\right]^{\rho} \quad (21)$$

or equivalently

$$\mathcal{K}_0(\boldsymbol{y}, C, \lambda, \rho) \geq \left[1 - \mathcal{X}^{\lambda \mathrm{wt}(\boldsymbol{c}_m)} + \sum_{m' \neq 0}\mathcal{X}^{\lambda \mathrm{wt}(\boldsymbol{c}_{m'})}\right]^{\rho}. \quad (22)$$

The latter expression indicates that the minimum value of $\mathcal{K}_0(\boldsymbol{y}, C, \lambda, \rho)$ is achieved when $\mathrm{wt}(\boldsymbol{c}_m)$ is minimum and not equal to 0, since $\mathcal{X} \leq 1$ and the inner sum of the right-hand side of (22) is constant. Thus, due to the equivalence of the distance and weight enumerating functions

$$\mathcal{K}_0(\boldsymbol{y}, C, \lambda, \rho) \geq \left[1 - \mathcal{X}^{\lambda d_{\min}} + \sum_{d=d_{\min}}^{n} S_d \mathcal{X}^{\lambda d}\right]^{\rho}. \quad (23)$$

Replacing (23) in the right-hand side of the first inequality of (8), and employing the DS2 technique with the arbitrary nonnegative function $G^m(\boldsymbol{y}) = \prod_{i=1}^{n} g(y_i)$, we arrive at the following theorem.

Theorem 5: Consider the transmission of a set of messages $\mathcal{M}$ through a binary linear (n,k) code C over a discrete memoryless output-symmetric channel. Given that the all-zero codeword $\boldsymbol{c}_0$ is transmitted, the ML word error decoding probability is upper-bounded for all $\lambda \geq 0, 0 \leq \rho \leq 1$ and any nonnegative function $g(y)$ by

$$P_{e|0} \leq \left[1 - \mathcal{X}^{\lambda d_{\min}} + \sum_{d=d_{\min}}^{n} S_d \mathcal{X}^{\lambda d}\right]^{-\rho} \cdot \left(\sum_{y}\Pr(y|0)g(y)\right)^{n(1-\rho)}$$

$$\cdot\left(\sum_{d=d_{\min}}^{n} S_d \left(\sum_y \Pr(y|0) g(y)^{1-1/\rho}\right)^{n-d} \cdot\left(\sum_y \Pr(y|0)^{1-\lambda}\Pr(y|1)^{\lambda} g(y)^{1-1/\rho}\right)^{d}\right)^{\rho}. \quad (24)$$

The proposed technique also covers those special cases of the DS2 bound, where the corresponding technique is applied to every term $P_{e|0}(d)$ of the overall union bound [2, eq. (4.45)]

$$P_{e|0} \leq \sum_{d=d_{\min}}^{n} P_{e|0}(d). \quad (25)$$

Specifically, for every subcode C_d, containing all codewords $\boldsymbol{c}_m$ of weight d

$$P_{e|0}(d) \leq \left(\frac{1}{1+(S_d-1)\mathcal{X}^{\lambda d}}\right)^{\rho} S_d^{\rho} \cdot\left(\sum_y \Pr(y|0) g(y)^{1-1/\rho}\right)^{(n-d)\rho} \cdot\left(\sum_y \Pr(y|0)^{1-\lambda}\Pr(y|1)^{\lambda} g(y)^{1-1/\rho}\right)^{d\rho} \quad (26)$$

since the minimum value of the left-hand side of (21) is considered over the erroneous decoding region of C_d.

For small values of $\mathcal{X}$, the lower bound (21) and hence the first factor of the right-hand side of (24) are close to one since the corresponding terms $S_d\mathcal{X}^d, d = d_{\min},\dots,n$ are small. Thus for high signal-to-noise ratio (SNR), the upper bound (24) is close to Gallager's upper bound (1). In low-SNR regimes, the impact of this factor becomes substantial. Illustrating examples are provided in Section IV.

Analogous results hold for the bit error probability analysis. More precisely we have

$$\mathcal{K}_0'(\boldsymbol{y}, C, U, \lambda, \rho) \geq \left[d(\boldsymbol{u}_m, \boldsymbol{u}_0)\left(1-\mathcal{X}^{\lambda \mathrm{wt}(\boldsymbol{c}_m)}\right) + \sum_{m'\neq 0} d(\boldsymbol{u}_{m'}, \boldsymbol{u}_0)\mathcal{X}^{\lambda \mathrm{wt}(\boldsymbol{c}_{m'})}\right]^{\rho} \quad (27)$$

and the minimum value achieved by $\mathcal{K}_0'(\boldsymbol{y}, C, U, \lambda, \rho)$, is

$$\mathcal{K}_0'(\boldsymbol{y}, C, U, \lambda, \rho) \geq \left[\left(\min_{\substack{m:\\ \mathrm{wt}(\boldsymbol{c}_m)=d_{\min}}} d(\boldsymbol{u}_m, \boldsymbol{u}_0)\left(1-\mathcal{X}^{\lambda d_{\min}}\right)\right) + \sum_{d=d_{\min}}^{n} S_d'\mathcal{X}^{\lambda d}\right]^{\rho} \quad (28)$$

with $S_d' = \sum_{w=1}^{k} A_{w,d}$ and $A_{w,d}$ the input–output weight enumerating function of the systematic code C.

B. Discrete Channels and Coset-Based Analysis

Since the channel is memoryless and output symmetric, each fraction in (19) becomes

$$\left(\frac{\Pr(\boldsymbol{y}|\boldsymbol{c}^{\dagger})}{\Pr(\boldsymbol{y}|\boldsymbol{c}_0)}\right)^{\lambda} = \left((q-1)\frac{1-p}{p}\right)^{\lambda\left(\mathrm{wt}(\boldsymbol{y})-\mathrm{wt}(\boldsymbol{y}-\boldsymbol{c}^{\dagger})\right)} \quad (29)$$

where $\boldsymbol{c}_0$ the all zero codeword. Given a specific $\boldsymbol{y}^* \in \boldsymbol{Y}_0^b$, in analogy to [9, eq. (33)], we define

$$\deg(\boldsymbol{y}^*|\boldsymbol{c}_0) = \left|\left\{\boldsymbol{c}_m \in C, m \neq 0 : \mathrm{wt}(\boldsymbol{y}^* - \boldsymbol{c}_m) \leq \mathrm{wt}(\boldsymbol{y}^*)\right\}\right|. \quad (30)$$

$\deg(\boldsymbol{y}^*|\boldsymbol{c}_0)$ is the number of codewords whose Hamming distance from $\boldsymbol{y}^*$ is lower than or equal to $\mathrm{wt}(\boldsymbol{y}^*)$. The corresponding ratio (29) for each of the above codewords is greater or equal to 1 so that

$$\left(\sum_{m'\neq 0}\left(\frac{\Pr(\boldsymbol{y}^*|\boldsymbol{c}_{m'})}{\Pr(\boldsymbol{y}^*|\boldsymbol{c}_0)}\right)^{\lambda}\right)^{\rho} \geq \deg(\boldsymbol{y}^*|\boldsymbol{c}_0)^{\rho}. \quad (31)$$

In analogy again to [9, eq. (43)], for every $\boldsymbol{y}^* \in \boldsymbol{Y}_0^b$

$$\deg(\boldsymbol{y}^*|\boldsymbol{c}_0) = \sum_{w=1}^{j_t} B_w^t - 1, \quad \left\lceil\frac{d_{\min}}{2}\right\rceil \leq j_t < n \quad (32)$$

where t denotes the coset of $\boldsymbol{y}^*$, $j_t = \mathrm{wt}(\boldsymbol{y}^*)$ and B_w^t is the number of words of weight w in the coset t. Contrary to [9, eq. (43)], the term $B_{j_t}^t$ contributes to the sum in the right-hand side of (32), since the inequality in (30) is not strict. Moreover, the absence of codeword $\boldsymbol{c}_0$ in the previous definition justifies reducing by 1 the aforementioned sum. Consequently, through (31) and (32)

$$\min_{\boldsymbol{y}^*\in\boldsymbol{Y}_0^b}\left(\sum_{m'\neq 0}\left(\frac{\Pr(\boldsymbol{y}^*|\boldsymbol{c}_{m'})}{\Pr(\boldsymbol{y}^*|\boldsymbol{c}_0)}\right)^{\lambda}\right)^{\rho} \geq \min_{t\in\mathcal{T}}\left(\min_{\lceil d_{\min}/2\rceil\leq j_t<n}\sum_{w=1}^{j_t} B_w^t - 1\right)^{\rho} \quad (33)$$

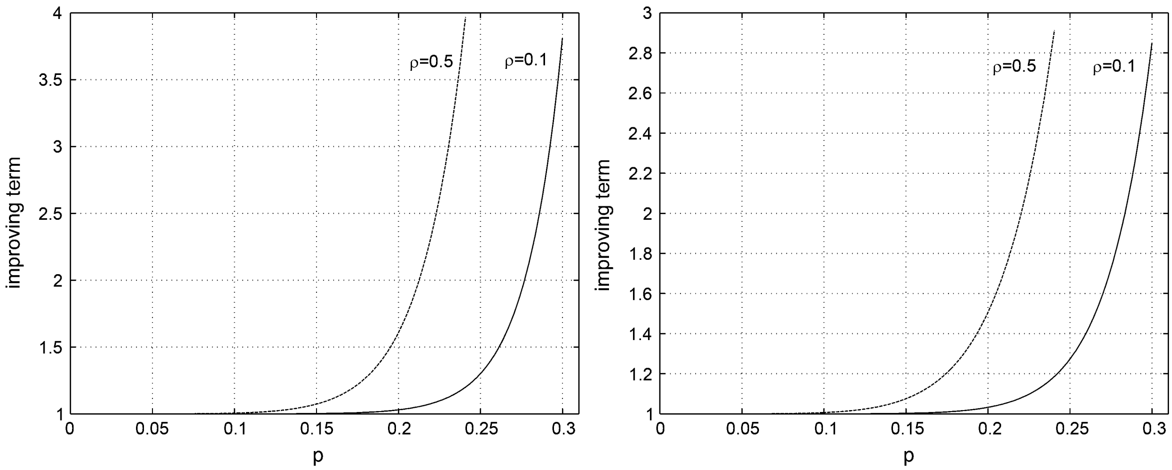
where $\mathcal{T}$ is the set of all cosets t of the code C.

Theorem 6: Under the assumptions of Theorem 5, the ML word error decoding probability is upper-bounded for all $\lambda \geq 0, 0 \leq \rho \leq 1$ and any nonnegative function $g(y)$ by

$$P_{e|0} \leq \left(\min_{t\in\mathcal{T}}\left(\min_{\lceil d_{\min}/2\rceil\leq j<n}\sum_{w=1}^{j} B_w^t - 1\right)^{\rho}\right)^{-1} \cdot\left(\sum_y \Pr(y|0) g(y)\right)^{n(1-\rho)} \cdot\left(\sum_{d=d_{\min}}^{n} S_d\left(\sum_y \Pr(y|0) g(y)^{1-1/\rho}\right)^{n-d} \cdot\left(\sum_y \Pr(y|0)^{1-\lambda}\Pr(y|1)^{\lambda} g(y)^{1-1/\rho}\right)^{d}\right)^{\rho}. \quad (34)$$

Example 1: Consider the perfect Hamming code of length 7 with its coset weight distribution depicted in Table I [10, p. 170, example (1)]. Since the minimum distance of the code is 3, all cosets with minimum weight at least $\lceil 1.5\rceil$ are examined. Then for $j_t = 2,\dots,7$

$$\min_{t\in\mathcal{T}}\left(\min_{2\leq j_t<7}\sum_{w=1}^{j_t} B_w^t - 1\right)^{\rho} = (3+1-1)^{\rho}. \quad (35)$$



(a) The improving term of the exrended (24, 12, 8) binary Golay code. (b) The improving term of the (32, 16, 8) Reed Muller code.

Fig. 1. The improving term (first term of the product (24)) of binary codes over the memoryless binary-symmetric channel with error transition probability p and tradeoff variable $\rho = 0.1, \rho = 0.5$, respectively.

TABLE I
THE COSET WEIGHT DISTRIBUTION OF THE PERFECT HAMMING CODE OF LENGTH 7

Coset leader Weight	Number of vectors of given weight 0	1	2	3	4	5	6	7	Number of cosets
0	1	0	0	7	7	0	0	1	1
1	0	1	3	4	4	3	1	0	7

The minimum value is achieved for $j_t = 2$, since for $j_t > 2$, the sum over $\mathcal{T}$ in the right-hand side of (33) increases. Actually, since the minimum distance of the code is an odd number, there will always exist a term in the left-hand side of (33) strictly greater than one.

Note that for a binary linear code C, the cardinality of each term $B_w^t \neq 0$, $w > 0$ appearing in every coset t stems from codewords $\boldsymbol{c}_m, m \neq 0$ with weight less than or equal to $2w$. Indeed, from (30) and the triangle inequality $\mathrm{wt}(\boldsymbol{y}^* - \boldsymbol{c}_m) + \mathrm{wt}(\boldsymbol{y}^*) \geq \mathrm{wt}(\boldsymbol{c}_m)$, it is deduced $2\mathrm{wt}(\boldsymbol{y}^*) \geq \mathrm{wt}(\boldsymbol{c}_m)$. Thus, if $\mathrm{wt}(\boldsymbol{y})^* = w$, it holds $\mathrm{wt}(\boldsymbol{c}_m) \leq 2w$.

The above argument is used to upper-bound the bit error decoding probability of a systematic code C. Let $\hat{t}$, $\hat{j}_{\hat{t}}$ denote the optimum values that minimize the right-hand side of (33). The codewords $\boldsymbol{c}_m, m \neq 0$ that contribute to the sum $\sum_{w=1}^{\hat{j}_{\hat{t}}} B_w^{\hat{t}} - 1$ have Hamming weight less than or equal to $2\hat{j}_{\hat{t}}$. Among these, only $\sum_{w=1}^{2\hat{j}_{\hat{t}}} A_{1,w}$ have information block weights equal to 1, while the remaining, $\sum_{w=1}^{\hat{j}_{\hat{t}}} B_w^{\hat{t}} - 1 - \sum_{w=1}^{2\hat{j}_{\hat{t}}} A_{1,w}$, have information block weights greater than or equal to 2. Thus

$$
\begin{aligned}
K_0'(\boldsymbol{y}, C, U, \lambda, \rho) &\geq \sum_{w=1}^{2\hat{j}_{\hat{t}}} A_{1,w} + 2\left(\sum_{w=1}^{\hat{j}_{\hat{t}}} B_w^{\hat{t}} - 1 - \sum_{w=1}^{2\hat{j}_{\hat{t}}} A_{1,w}\right) \\
&= 2\sum_{w=1}^{\hat{j}_{\hat{t}}} B_w^{\hat{t}} - 2 - \sum_{w=1}^{2\hat{j}_{\hat{t}}} A_{1,w}. \qquad (36)
\end{aligned}
$$

Consequently, due to (17), (36), and the DS2 technique the following statement holds.

Theorem 7: Under the assumptions of Theorem 5, the ML bit error decoding probability is upper-bounded for all $\lambda \geq 0, 0 \leq \rho \leq 1$ and any nonnegative function $g(y)$ by

$$
\begin{aligned}
P_{b|0} \leq & \left(\min_{t \in \mathcal{T}} \left(\min_{\lceil d_{\min}/2 \rceil \leq j_t < n} 2\sum_{w=1}^{j_t} B_w^t - 2 - \sum_{w=1}^{2j_t} A_{1,w}\right)^{\rho}\right)^{-1} \\
& \cdot \left(\sum_y \Pr(y|0) g(y)\right)^{n(1-\rho)} \\
& \cdot \left(\sum_{d=d_{\min}}^{n} S_d' \left(\sum_y \Pr(y|0) g(y)^{1-1/\rho}\right)^{n-d}\right. \\
& \left. \cdot \left(\sum_y \Pr(y|0)^{1-\lambda} \Pr(y|1)^{\lambda} g(y)^{1-1/\rho}\right)^{d}\right)^{\rho}. \qquad (37)
\end{aligned}
$$

Note that if the codewords with information block weights equal to 1 are removed, a nonlinear code with lower bit error decoding probability may result.

IV. APPLICATIONS AND CONCLUSION

The transmission of coded information over a memoryless binary input and output-symmetric channel is considered. The extended $(24, 12, 8)$ binary Golay code and the $(32, 16, 8)$ Reed–Muller code are utilized. The coset weight distributions of these codes are given in [11, Table 8.4] and [11, Table 11.4], respectively. For the Golay code, the first term of the product in (24) is calculated for error transition probabilities $p \leq 0.3$ and is depicted in Fig. 1(a). For $p > 0.15$, the corresponding upper bound in (24) is only slightly better than the Gallager bound. On the other hand, the minimum in the right-hand side of (33) occurs for the coset $\hat{t}$ with $\hat{j}_{\hat{t}} = \lceil d_{\min}/2 \rceil = 4$ and is equal to $(B_4^{\hat{t}})^{\rho} = 5^{\rho}$. Thus, the new upper bound (34) is $1/5^{\rho}$ tighter than the DS2 bound. Regarding the bit error probability decoding for the Golay code, only $A_{1,8} = 3$ codewords of weight up to $2\hat{j}_{\hat{t}} = 8$ have information block weight equal to 1 while $B_4^{\hat{t}} - A_{1,8} = 2$ have information block weight greater than 1. Thus, due to (36), the new upper bound (37) is $1/7^{\rho}$ tighter than the DS2 bound. For the Reed–Muller code, in the calculation of the right-hand side of (33), the family with coset leader weight equal to 1 is not

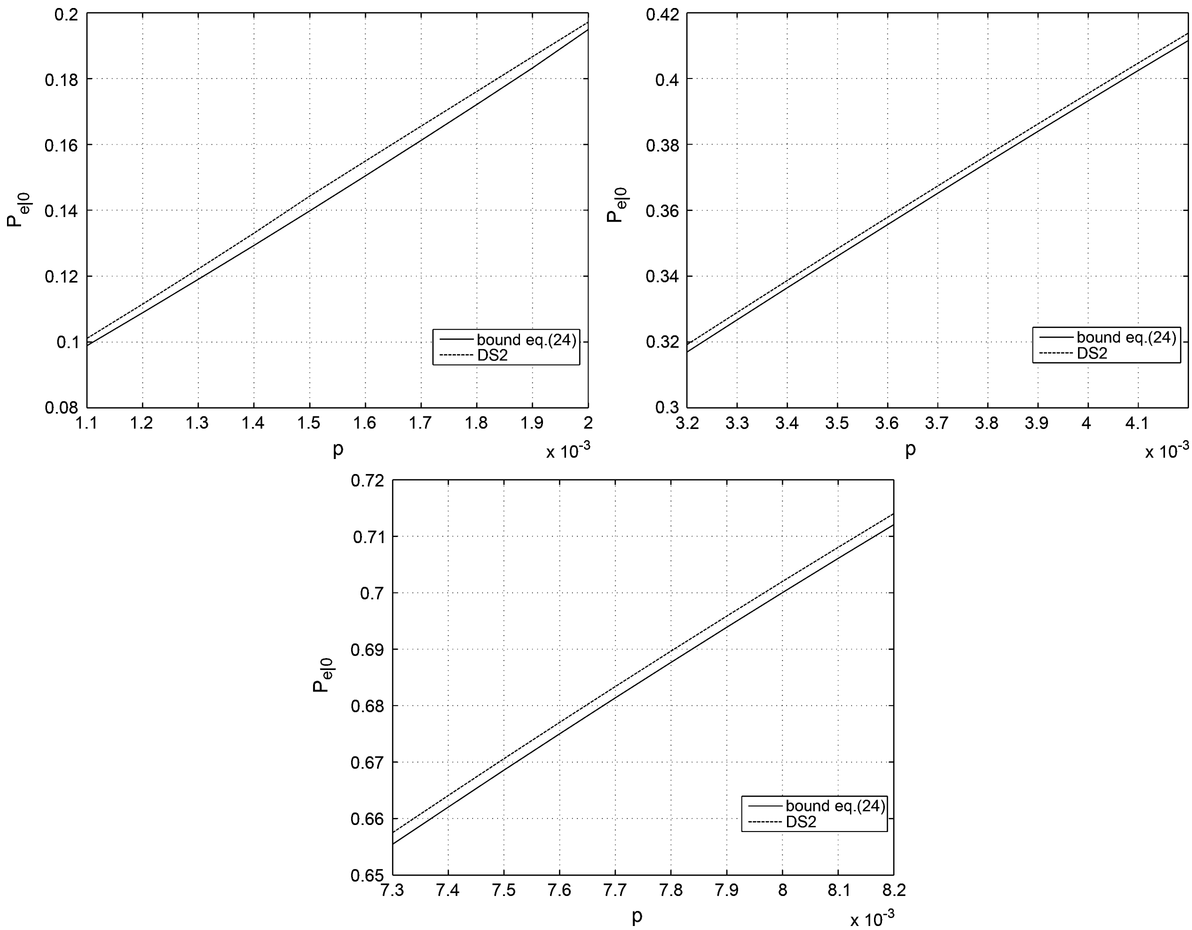


Fig. 2. The new upper bound (24) and the classical DS2 bound (3) on the error probability $P_{e|0}$, with un-normalized tilting measure (38), for the transmission of the $(60, 38)$ block code over a binary-symmetric channel with error transition probability p.

taken into consideration, due to the restriction imposed by the set definition (30). Hence, the minimum value of the inner term on the right-hand side of (33) is equal to $(2-1)^\rho = 1$ and thus no improvement in Gallager's first upper bound and its DS2 variations results through the specific technique for the Reed–Muller code. The first term of the product in (24) is calculated again for error transition probabilities $p \leq 0.3$ and only slight improvements are yielded, as depicted in Fig. 1(b).

Finally, we consider the transmission of an $(60, 38)$ block code over a binary-symmetric channel with error transition probability p. The previous code is constructed by terminating, in depth 20, the $R = 2/3$ convolutional code tabulated in [12], with generator matrix

$$G = \begin{pmatrix} 2 & 0 & 1 \\ 2 & 3 & 0 \end{pmatrix}.$$

The proposed bound of (24) and the classical DS2 bound (3) are minimized over $\lambda > 0, 0 \leq \rho \leq 1$, with un-normalized tilting measure $g(y)$

$$g(y) = \left(\frac{1}{2}\Pr(y|0)^\lambda + \frac{1}{2}\Pr(y|1)^\lambda\right)^p \Pr(y|0)^{-\lambda\rho}. \quad (38)$$

The full weight distribution of the code is taken into consideration in the previous minimization procedure. The two upper bounds are illustrated in Fig. 2, for regions of p where the term in the right-hand side of (23) contributes most. In all other cases, the two bounds are essentially the same.

We conclude that a desirable design characteristic of a specific code for discrete memoryless channels is that the first few terms in the coset weight distribution, with weight greater than or equal to one half the code's minimum distance, are as large as possible. This aspect is closely related to the list decoding concept, since when an error occurs all the codewords in the list fail the decoding rule. Finally, the amenability of the new upper bound to mismatched decoding rules, especially for continuous output channels, and its implications for new efficient code designs is under study.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewer for his helpful comments. They also wish to thank Alexandros Katsiotis for his assistance in the experimental part.

REFERENCES

[1] R. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inf. Theory*, vol. IT-11, no. 1, pp. 3–18, Jan. 1965.

[2] I. Sason and S. Shamai (Shitz), "Performance analysis of linear codes under maximum-likelihood decoding: A tutorial," *Foundations and Trends in Communications and Information Theory*, vol. 3, 2006.

[3] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near shannon limit error-correcting coding and decoding," in *Proc. IEEE Int. Conf. Communications (ICC'93)*, Geneva, Switzerland, May 1993, pp. 1064–1070.

[4] R. Gallager, *Low-Density Parity-Check Codes*. Cambridge, MA: MIT Press, 1963.

[5] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley-Interscience, 2006.
[6] T. Duman and M. Salehi, "New performance bounds for turbo codes," *IEEE Trans. Commun.*, vol. 46, no. 6, pp. 717–723, Jun. 1998.
[7] D. Divsalar, "A simple tight bound on error probability of block codes with application to turbo codes," *JPL, TMO Progr. Rep.*, pp. 42–139, Nov. 1999.
[8] E. Agrell, "Voronoi regions for binary linear block codes," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 310–316, Jan. 1996.
[9] A. Cohen and N. Merhav, "Lower bounds on the error probability of block codes based on improvements on de caen's inequality," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 290–310, Feb. 2004.
[10] F. MacWilliams and N. Sloane, *The Theory of Error-Correcting Codes*. Amsterdam, The Netherlands: North-Holland, 1983.
[11] W. Huffman and V. Pless, *Fundamentals of Error Correcting Codes*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
[12] H. Tang, M. Lin, and B. Uchoa-Filho, "Minimal trellis modules and equivalent convolutional codes," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3738–3746, Aug. 2006.

Kostis Xenoulis (S'08) received the B.Sc. degree in informatics and telecommunications and the M.Sc. degree in signal processing for telecommunications and multimedia from the University of Athens, Athens, Greece, in 2003 and 2005, respectively.

He is currently working towards the Ph.D. degree in the Department of Informatics and Telecommunications at the University of Athens. His research interests are in the area of information theory.

Nicholas Kalouptsidis (M'82–SM'85) was born in Athens, Greece, on September 13, 1951. He received the B.Sc. degree in mathematics (with highest honors) from the University of Athens, Athens, Greece, in 1973 and the M.S. and Ph.D. degrees in systems science and mathematics from Washington University, St. Louis, MO, in 1975 and 1976, respectively.

He has held visiting positions at Washington University, St. Louis, MO; Politecnico di Torino, Turin, Italy; Northeastern University, Boston, MA; and CNET Lannion, France. He has been an Associate Professor and Professor with the Department of Physics, University of Athens. In Fall 1998, he was a Clyde Chair Professor with the School of Engineering, University of Utah, Salt Lake City. In Spring 2008, he was a Visiting Scholar at Harvard University, Cambridge, MA. He is currently a Professor with the Department of Informatics and Telecommunications, University of Athens. He is the author of the textbook *Signal Processing Systems: Theory and Design* (New York: Wiley, 1997) and coeditor, with S. Theodoridis, of the book *Adaptive System Identification and Signal Processing Algorithms* (Englewood Cliffs, NJ: Prentice-Hall, 1993). His research interests are in system theory and signal processing.