List permutation invariant linear codes: Theory and Applications

Kostis Xenoulis

***Abstract*—The class of q-ary list permutation invariant linear codes is introduced in the current work along with probabilistic arguments that validate their existence when certain conditions are met. The specific class of codes is characterized by an upper bound that is tighter than the generalized Shulman-Feder bound and relies on the distance of the codes' weight distribution to the binomial (multinomial respectively) one. The bound applies to cases where a code from the proposed class is transmitted over a q-ary output symmetric discrete memoryless channel and list decoding with fixed list size is performed at the output. In the binary case, the new upper bounding technique allows the discovery of list permutation invariant codes whose upper bound coincides with sphere-packing exponent. Furthermore, the proposed technique motivates the introduction of a new class of upper bounds for general q-ary linear codes whose members are at least as tight as the DS2 bound as well as all its variations for the discrete channels treated in the current work.**

***Index Terms*—Discrete symmetric channels, double exponential function, list decoding, permutation invariance, reliability function.**

I. Introduction

The development of efficient and simplified bounding techniques which apply to structured families of codes constitutes a major research area in the field of coded communications. Their importance lies in their ability to designate characteristics under which the proposed codes achieve low error decoding probability with rates close to the capacity of the considered channels. A vast majority of bounding techniques have been constructed for structured ensembles of codes over discrete and continuous channels, as it is reported in the literature [1]. Their usefulness relies among others on their ability to apply to specific families as well as to fully random ensembles of codes. A classical representative of the class of these techniques, which applies to both categories of codes, is the generalized second version of the Duman-Salehi bound (DS2) [2, Section II.B]. For specific codes, many reported bounds such as Shulman-Feder bound (SF) [3] for discrete memoryless channels and Divsalar bound [4] for binary input additive Gaussian channels, consist special cases of the DS2 bound. The latter can be applied to the fully random ensemble of codes and provide Gallager's random coding theorem [5].

An interesting class of bounding techniques for the error decoding probability of specific codes is the one which enjoys random coding characteristics, such as the SF bound and its generalized version, reported in [2, Appendix A]. The latter relates the coding exponent of nonrandom codes to the exponent deduced from the random coding analysis of Gallager [5]. Specifically, through the construction of an artificial ensemble of codes from a single block code, the difference between the error exponent of the specific code and the exponent of the random coding ensemble under maximum likelihood (ML) decoding is revealed to be equal to a constant term. This term depends only on the weight spectrum of the code and its relation to the binomial (multinomial) distribution and not on the channel's characteristics. The importance of the presented bounds lies on their close similarity to the sphere packing lower bound [6] on the error decoding probability and its tightest versions for finite blocklengths, presented in [7] and [8] respectively.

The purpose of the current work is to provide new tight upper bounds for the fixed size list error decoding probability, both in the spirit of the SF bound and its generalized version, that are able to approach the sphere-packing lower bound. Exponential error bounds and random coding exponents for list error decoding with variable list size have been established in [9] for the class of randomly constructed codes. Similar results under fixed-size list decoding were presently recently in [10] for the class of fixed-composition codes. Regarding specific codes, error performance bounds under fixed-size list-decoding are provided in [11] and [12]. In [11] the bounds specialize to linear codes and continuous output channels and rely on the distances between the codewords on the list. On the other hand, [12] focuses on the derivation of Gallager-type bounds both under variable and fixed size list decoding. The presented bounds are message independent for general-symmetric channels and apply to both random and structured linear block codes (or code ensembles).

Regarding the case of (structured) code ensembles, the bounds presented in the current work are deployed by replacing weighted sums of likelihood ratios in the DS2 technique as well as in the proof of the random coding theorem by double exponential functionals. We perform so in order to exploit the concave behavior of the double exponential function in the error decoding region of a code under fixed size list decoding and combine it with the random coding technique. To proceed further, the class of L-list permutation invariant linear codes is introduced. It is revealed that every member of the class enjoys the same list error decoding probability for the transmission of coded information over discrete memoryless symmetric channels. The specific approach constitutes a new design paradigm where the need for bounding techniques, that lead to tight closed-form upper bounds, enables the definition of a new class of codes. For a list permutation invariant linear

This research has been co-financed by the European Union (European Social Fund-ESF) and Greek national funds through the Operational Program Education and Lifelong Learning of the National Strategic Reference Framework (NSRF) - Research Funding Program: Aristeia I-1115.

The author is with the Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Athens, Greece 15784, e-mail: kxen@di.uoa.gr

code, the proposed bounding technique leads to tight upper bounds in relation to SF bound and its generalized version while it still relies on the similarity of the code's weight distribution to the binomial one. The analytic construction of members belonging to the previous class of codes constitutes a heavy computational process since it relies on the satisfaction of certain properties of a code's distance distribution matrix. Nevertheless, the probabilistic method [13] and specifically Lovasz's Local Lemma allows the proof of existence of such a class of codes and renders them as a useful tool in the calculation of the reliability function of discrete memoryless channels for certain coding rates.

The double exponential technique, as described in the previous paragraph, is relative new and has appeared in [14]. In general, double exponential bounds have appeared in the analysis of error exponents for channels with feedback, such as in [15] and references therein, but are not related to the presented technique. The latter can also be applied to general q-ary linear codes leading to a tighter version of the DS2 bound. Nevertheless, it violates the need for simplified bounds since it requires the knowledge of fine details of a code's characteristics, such as the full coset weight distribution. The previous obstacle is circumvented by noting that the key ingredient of the DS2 bound is a weighted average of the sum of likelihood ratios of the code's codewords with respect to the transmitted one, and that this sum can be replaced with respective sums of horizontally flip sigmoid functions. The latter manage to smooth the effect of the erroneous decoding region in the calculation of the list error decoding probability and thus achieve a tradeoff between the desirable characteristics of bounding techniques. Namely, the presented bounds are tighter than the DS2 bound while their evaluation does not require any extra knowledge in comparison to the DS2 bound.

The present paper extends the work in [14] with the introduction of the class of q-ary L-list permutation invariant linear codes and the analysis of the error performance of its members under list decoding for transmission over q-ary discrete symmetric memoryless channels. Towards the extension of list permutation invariant codes from binary to q-ary fields and in comparison to [14], a new definition of L-list permutation invariant codes is provided in order to address a broader class of candidate codes. Even though the extension to the general q-ary case is straightforward, the derivation of an analytic error bound expression is rather complicated. In what follows, Section II specifies the properties underpinning the class of L-list permutation invariant linear codes and presents a relaxed condition under which such codes appear to exist. Under this condition, subsection II-A validates the existence of permutation invariant codes and enables the analysis of their list-error decoding probability, leading to the upper bounds presented in Sections III and III-A respectively. The class of permutation invariant codes is utilized in Section IV to provide the reliability function of binary symmetric channels for certain coding rates. Finally, a new class of bounding techniques is presented in Section V for general q-ary linear codes, while Section VI concludes the analysis by numerically comparing the developed bounds for the new class of codes.

II. Permutation Invariance

Let the transmission of an arbitrary set of messages $\mathcal{M}$, with cardinality M, over a discrete memoryless q-ary output symmetric channel, with the use of an (N,R) linear code $\mathcal{C}$ defined over the finite field $F_q \triangleq \{0,1,\ldots,q-1\}$. Specifically, each message $m \in [0, M-1]$ with $M = q^{NR}$, is encoded onto an N length codeword $\mathbf{c}_m \in F_q^N$ of $\mathcal{C}$, where $\mathbf{c}_0$ is the all-zero codeword, and is transmitted over the symmetric channel described by the transition probability $\Pr(\mathbf{y}|\mathbf{c}_m)$, where $\mathbf{y} \in F_q^N$ the N length sequence at the output of the channel. During the transmission, each symbol of a codeword remains the same with probability $1-p$, while it is changed to another of the remaining $q-1$ symbols with probability $p/(q-1)$. In the analysis that follows, we denote by $d_{\min}$ the minimum Hamming weight of the linear code $\mathcal{C}$ and by $S_{l_1,\ldots,l_{q-1}}$ the number of codewords of $\mathcal{C}$ having a fixed type [16, Ch. 11] with l_i occurrences of symbol i. Furthermore, we denote the weight distribution of a linear code $\mathcal{C}$ over F_q as the sequence $\{S_k\}_{k=d_{\min}}^N$ where

$$S_k \triangleq \sum_{l_1+\ldots+l_{q-1}=k} S_{l_1,\ldots,l_{q-1}}. \tag{1}$$

The translate set $\mathbf{y}-\mathcal{C} \triangleq \{\mathbf{y}-\mathbf{c}, \mathbf{c}\in\mathcal{C}\}$ is called a coset of the linear code $\mathcal{C}$, while the vector with the smallest Hamming weight in the aforementioned set is called a coset leader. The weights of the vectors in the coset $\mathbf{y}-\mathcal{C}$ define the weight distribution of the considered coset, while the largest weight of a coset leader equals the covering radius of $\mathcal{C}$, denoted by $\rho(\mathcal{C})$. The relative radius of $\mathcal{C}$ is defined respectively as $\rho_{\mathrm{r}}(\mathcal{C}) \triangleq \rho(\mathcal{C})/N$.

Definition 1: The distance distribution matrix $\mathbf{\Gamma}$ of a linear code $\mathcal{C}$ is defined as the $K \times (N+1)$ matrix having as rows all the possible K different coset weight distributions. The first row of $\mathbf{\Gamma}$ defines the weight distribution of $\mathcal{C}$, while all other rows are identified by the weight of the coset leader (coset weight). If the weight distribution of the coset of the received vector $\mathbf{y}$ belongs to the i-th row of the matrix, then $\mathbf{y}$ has $\sum_{k=0}^t \mathbf{\Gamma}_{i,k}$ codewords within Hamming distance t. Furthermore, the j-th column of $\mathbf{\Gamma}$, $j\in[0,N]$, denotes the number of codewords a vector $\mathbf{y}$ of Hamming weight i can have at exact Hamming distance j.

Considering list decoding is performed at the output of the channel with list size L, the conditional error decoding probability given the transmission of message 0 is

$$P_{e|0,\mathcal{C}}^L = \sum_{\mathbf{y}\in\mathbf{Y}_{0,\mathcal{C}}^L} \Pr(\mathbf{y}|\mathbf{c}_0,\mathcal{C}) \tag{2}$$

where

$$\mathbf{Y}_{0,\mathcal{C}}^L \triangleq \Big\{\mathbf{y}\in F_q^N : \exists\{l_i\}_{i=1}^L \\ \text{with } \Pr(\mathbf{y}|\mathbf{c}_{l_i},\mathcal{C}) \geq \Pr(\mathbf{y}|\mathbf{c}_0,\mathcal{C}),\ l_i \neq 0\Big\}. \tag{3}$$

Since the channel is memoryless and output symmetric, the error decoding region $\mathbf{Y}_{0,\mathcal{C}}^L$ is equivalently expressed as

$$\mathbf{Y}_{0,\mathcal{C}}^L = \Big\{\mathbf{y}\in F_q^N : \exists\{l_i\}_{i=1}^L \\ \text{with } \mathrm{wt}(\mathbf{y}-\mathbf{c}_{l_i}) \leq \mathrm{wt}(\mathbf{y}),\ l_i\neq 0\Big\} \tag{4}$$

where wt$(\mathbf{y})$ denotes the Hamming weight of a q-ary vector $\mathbf{y}$, with the respective algebra defined over the finite field F_q. Moreover, if for $\lambda, \rho \geq 0$, we define as in [14]

$$\Omega(\mathbf{y}, \mathcal{C}, \lambda, \rho) \triangleq \frac{L^{\rho} \mathrm{Pr}(\mathbf{y}|\mathbf{c}_0)^{\lambda\rho}}{\left(\sum_{m\neq 0} \mathrm{Pr}(\mathbf{y}|\mathbf{c}_m)^{\lambda}\right)^{\rho}} \tag{5}$$

then, for every $\mathbf{y} \in \mathbf{Y}^L_{0,\mathcal{C}}$ it holds

$$1 \leq f_{\text{de}}\left(\Omega(\mathbf{y}, \mathcal{C}, \lambda, \rho)\right), \quad f_{\text{de}}(x) \triangleq \exp\left(1 - \exp\left(x - 1\right)\right) \tag{6}$$

and the error decoding probability $P^L_{e|0,\mathcal{C}}$ in (2) is upper bounded for any $\lambda, \rho \geq 0$ as

$$P^L_{e|0,\mathcal{C}} \leq \sum_{\mathbf{y}\in \mathbf{Y}^L_{0,\mathcal{C}}} \mathrm{Pr}(\mathbf{y}|\mathbf{c}_0, \mathcal{C}) f_{\text{de}}(\Omega(\mathbf{y}, \mathcal{C}, \lambda, \rho)). \tag{7}$$

It is noted that $\Omega(\mathbf{y}, \mathcal{C}, \lambda, \rho)$ is exactly the inverse of the expression used to bound the indicator function $\phi_m(\mathbf{y})$ in [5, Eqs. (4),(5)]. If we proceed by extending the above sum over the set F_q^N, then we cannot easily exploit the concavity of the double exponential function. Thus we cannot average over the random ensemble of codes and employ Jensen's inequality to produce a random coding exponent. This is due to the fact that $f_{\text{de}}(\Omega(\mathbf{y}, \mathcal{C}, \lambda, \rho))$ is concave with respect to $\Omega(\mathbf{y}, \mathcal{C}, \lambda, \rho)$ only for $\mathbf{y} \in \mathbf{Y}^L_{0,\mathcal{C}}$. To surpass the previous obstacle and take advantage of the special form of (7), we seek for a family of linear codes where each code has the same error decoding region $\mathbf{Y}^L_{0,\mathcal{C}}$. This motivates us to define the class of L-list permutation invariant linear codes.

Definition 2: An (N, R) linear code $\mathcal{C}$ with distance distribution matrix $\mathbf{\Gamma}$ is called L-list permutation invariant if both the following properties are satisfied:

$\mathcal{L}_1$: there exists a positive integer $w_{opt} \in [\lceil d_{\min}/2\rceil, N/2)$ such that

$$\min_{\kappa\in[1,K]} \sum_{w=1}^{w_{opt}} \mathbf{\Gamma}_{\kappa,w_{opt}} \geq L + 1 \tag{8}$$

and[1]

$$\max_{\kappa\in[1,K]} \sum_{w=1}^{w_{opt}-1} \mathbf{\Gamma}_{k,w} \leq L. \tag{9}$$

$\mathcal{L}_2$: For all $\kappa \in [1, K]$ there exists a $w^{\mathcal{L}}_{\kappa} > w_{opt}$ such that $\mathbf{\Gamma}_{\kappa,w_{opt}+1} = \ldots = \mathbf{\Gamma}_{\kappa,w^{\mathcal{L}}_{\kappa}-1} = 0$ and $\mathbf{\Gamma}_{\kappa,w^{\mathcal{L}}_{\kappa}} \geq L + 1$.

In case $L = 1$, we call the specific code ML permutation invariant.

Example 1: As an example, we consider the ternary $[13, 7, 5]$ quadratic-residue code of [17] with distance distribution matrix depicted in Table I. If we select $w_{opt} = \lceil d_{\min}/2\rceil = 3$, then for $L = 1$, property $\mathcal{L}_1$ is satisfied, since the sum of the columns before the 3-rd one is lower than or equal to 1. Furthermore, property $\mathcal{L}_2$ is also satisfied since for every row, in columns 4 and 5 there is an element greater than 2. Consequently, the specific code is ML permutation invariant. With similar arguments, selecting $w_{opt} = 4$ let us deduce that the specific code is also 6-list or 7-list or 8-list or 9-list or 10-list permutation invariant.

[1]Relation $\max_{\kappa\in[1,K],w<w_{opt}} \mathbf{\Gamma}_{k,w} \leq L$ was used in [14, Def. 2] instead of (9).

TABLE I
DISTANCE DISTRIBUTION MATRIX $\mathbf{\Gamma}$ OF THE TERNARY $[13, 7, 5]$ QUADRATIC-RESIDUE CODE OF [17, TABLE I]

0	1	2	3	4	5	6	7	8
1					78	182	286	390
	1			15	57	167	295	432
		1	4	11	54	164	304	435
		1	2	16	57	146	319	429
		1	2	14	64	144	296	469
		1	2	14	62	151	294	446
			5	15	50	163	294	446
			4	16	56	154	288	461
			3	20	55	138	311	463
			2	23	54	131	328	438
			4	18	51	149	313	444
			4	16	58	147	290	484

Remark 1: Property $\mathcal{L}_1$ in fact guarantees that if a vector $\mathbf{y}$ of Hamming weight $w_{opt} \geq \lceil d_{\min}/2\rceil$, which causes an L-list decoding error, is permuted then the permuted vector still causes an L-list decoding error. Namely, consider a vector $\mathbf{y}$ of weight w_{opt} that has $\mathbf{\Gamma}_{\kappa,w_{opt}} \neq 0$ codewords at exact Hamming distance w_{opt}. Then, $\mathbf{y}$ will have $\sum_{w=1}^{w_{opt}} \mathbf{\Gamma}_{\kappa,w}$ codewords within distance w_{opt}. Consider now a permuted vector $\mathbf{y}'$ of $\mathbf{y}$ and suppose that it has $\mathbf{\Gamma}_{\kappa',w_{opt}} \neq 0$ codewords at exact distance w_{opt} where $\kappa' \neq \kappa$. Then according to (8), the permuted vector of $\mathbf{y}$ will still have $\sum_{w=1}^{w_{opt}} \mathbf{\Gamma}_{\kappa',w} \geq L+1$ codewords within Hamming distance L. Furthermore, the specific property in (9) guarantees that all vectors of Hamming weight lower than w_{opt} do not cause an L-list decoding error. Regarding property $\mathcal{L}_2$, the latter ensures that all vectors $\mathbf{y}$ of Hamming weight greater than w_{opt} do cause an error. Indeed, under the validity of (8), in every row of the matrix $\mathbf{\Gamma}$, every sum of the elements of the respective row up to w for every $w > w_{opt}$ will be greater than $L + 1$.

As in [14], from an L-list permutation invariant linear code $\mathcal{C}$ we create an ensemble of linear codes $\mathcal{E}$ by considering all possible position permutation codes $\mathcal{PC}$. A $\mathcal{PC}$ code is constructed by multiplying every codeword of $\mathcal{C}$ with a $N \times N$ position permutation matrix $\mathcal{P}$ which has a single 1 in every row and column and is orthogonal, $\mathcal{P}\mathcal{P}^T = \mathcal{P}^T\mathcal{P} = I$, where I is the $N \times N$ identity matrix and $\mathcal{P}^T$ the transpose matrix of $\mathcal{P}$. The permutation ensemble $\mathcal{E}$ is utilized in order to prove the following property for an L-list permutation invariant linear code.

Lemma 1: For an (N, R) linear code $\mathcal{C}$ that is L-list permutation invariant, all codes in the position permutation ensemble $\mathcal{E}$ have the same error decoding region $\mathbf{Y}^{\mathcal{L}}_0$.

Proof: The proof of the lemma follows the same line as in the proof of [14, Lemma 1]. In particular, suppose that in the original L-list permutation invariant code $\mathcal{C}$ the received vector is $\mathbf{y} \in \mathbf{Y}^{\mathcal{L}}_{0,\mathcal{C}}$ where the weight distribution of the unique coset $\mathbf{y} - \mathcal{C}$ is $\{\mathbf{\Gamma}_{\nu,w}\}_{w=0}^N$. Due to the restriction imposed by condition $\mathfrak{L}_1$, it must be wt$(\mathbf{y}) \geq w_{opt}$. Indeed, if wt$(\mathbf{y}) < w_{opt}$, then according to (9) $\mathbf{y}$ will have at most $L - 1$ codewords, not including the all zeros codeword $\mathbf{c}_0$, within Hamming distance wt$(\mathbf{y})$. Thus this $\mathbf{y}$ will not cause a list decoding error.

Let such a received vector $\mathbf{y} \in \mathbf{Y}_{0,\mathcal{C}}^{\mathcal{L}}$ with $\mathrm{wt}(\mathbf{y}) = w_{opt}$. Then $\mathbf{y}$ has $\sum_{k=1}^{w_{opt}} \mathbf{\Gamma}_{\nu,k} - 1 \geq L$ codeword(s), different from $\mathbf{c}_0$, within Hamming distance w_{opt}. Let, for any permutation matrix $\mathcal{P}$, the corresponding permuted vector $\mathcal{P}^T\mathbf{y}$ of $\mathbf{y} \in \mathbf{Y}_{0,\mathcal{C}}^{\mathcal{L}}$. Vector $\mathcal{P}^T\mathbf{y}$ is also of weight w_{opt} while the weight distribution of coset $\mathcal{P}^T\mathbf{y}-\mathcal{C}$ is denoted by $\{\mathbf{\Gamma}_{\xi,w}\}_{w=0}^N$, where ξ does not necessarily equal ν. Since L is selected as the minimum sum over all rows of $\mathbf{\Gamma}$ up to the w_{opt}-th column minus 1 and $\sum_{w=1}^{w_{opt}-1} \mathbf{\Gamma}_{\kappa,w} \leq L$ for all $\kappa \in [1, K]$ (property $\mathfrak{L}_1$), vector $\mathcal{P}^T\mathbf{y}$ also has L codewords $\{\mathbf{c}_{l_i}\}_{i=1}^L$ of $\mathcal{C}$, apart from the all-zeros codeword $\mathbf{c}_0$, within distance $\mathrm{wt}(\mathbf{y}) = w_{opt}$, i.e. $\mathrm{wt}\left(\mathcal{P}^T\mathbf{y} - \mathbf{c}_{l_i}\right) \leq \mathrm{wt}(\mathbf{y}), i \in [1, L]$. Thus $\mathcal{P}^T\mathbf{y} \in \mathbf{Y}_{0,\mathcal{C}}^{\mathcal{L}}$. Moreover, according to the previous inequality it is noted

$$\begin{aligned}\mathrm{wt}(\mathbf{y}) \geq \mathrm{wt}\left(\mathcal{P}^T\mathbf{y} - \mathbf{c}_{l_i}\right) &= \mathrm{wt}\left(\mathcal{P}\left(\mathcal{P}^T\mathbf{y} - \mathbf{c}_{l_i}\right)\right) \\ &= \mathrm{wt}\left(\mathbf{y} - \mathcal{P}\mathbf{c}_{l_i}\right). \end{aligned} \tag{10}$$

The inequality presented in (10) indicates that $\mathbf{y}$ also has L codewords $\{\mathbf{c}_{l_i}\}_{i=1}^L$, apart from $\mathbf{c}_0$, of the permuted code $\mathcal{PC}$ within distance $\mathrm{wt}(\mathbf{y})$. Hence, $\mathbf{y} \in \mathbf{Y}_{0,\mathcal{C}}^{\mathcal{L}}$ causes a list decoding error in $\mathcal{PC}$.

Suppose now that $\mathbf{y} \in \mathbf{Y}_{0,\mathcal{C}}^{\mathcal{L}}$ with weight $\mathrm{wt}(\mathbf{y}) = w_\nu^{\mathcal{L}} \in (w_{\mathrm{opt}}, w_\kappa^{\mathcal{L}}]$, for some $\nu \in [1, K]$, and consider the weight distribution of the corresponding coset $\mathbf{y}-\mathcal{C}$, $\{\mathbf{\Gamma}_{\nu,w}\}_{w=0}^N$. For any permutation matrix $\mathcal{P}$, the weight distribution of coset $\mathcal{P}^T\mathbf{y} - \mathcal{C}$ is denoted by $\{\mathbf{\Gamma}_{\xi,w}\}_{w=0}^N$, where again ξ is not necessarily equal to ν. Then, due to property $\mathfrak{L}_2$, vector $\mathcal{P}^T\mathbf{y}$ has $\mathbf{\Gamma}_{\xi,w_\xi^{\mathcal{L}}} - 1 \geq L$ codewords of $\mathcal{C}$, apart from $\mathbf{c}_0$, at Hamming distance $w_\xi^{\mathcal{L}} \leq w_\nu^{\mathcal{L}}$. Equality holds in the latter relation in case cosets $\mathbf{y}-\mathcal{C}$ and $\mathcal{P}^T\mathbf{y}-\mathcal{C}$ have the same weight distribution. Consequently, $\mathcal{P}^T\mathbf{y} \in \mathbf{Y}_{0,\mathcal{C}}^{\mathcal{L}}$, while the weight transformation (10) guarantees that vector $\mathbf{y}$ also causes a list decoding error in the permuted code $\mathcal{PC}$. Finally, in all other cases where $\mathbf{y} \in \mathbf{Y}_{0,\mathcal{C}}^{\mathcal{L}}$ and $\mathrm{wt}(\mathbf{y}) > w_\kappa^{\mathcal{L}}$ for all $\kappa \in [1, K]$, vector $\mathcal{P}^T\mathbf{y}$ always has $\min_{\kappa\in[1,K]} \sum_{w=1}^{w_\kappa^{\mathcal{L}}} \mathbf{\Gamma}_{\kappa,w} - 1 \geq L$ codewords of $\mathcal{C}$ within distance $\mathrm{wt}(\mathbf{y})$ and thus $\mathcal{P}^T\mathbf{y} \in \mathbf{Y}_{0,\mathcal{C}}^{\mathcal{L}}$. ∎

The specific property stated in Lemma 1 constitutes the main reason for defining the class of L-list permutation invariant codes. In particular, taking the average of both sides of (7) over the permutation ensemble $\mathcal{E}$ we manage to change the order of summation with respect to $\mathbf{Y}_{0,\mathcal{C}}^L$ and $\mathcal{E}$. Since the double exponential function is concave over the subset $\mathbf{Y}_{0,\mathcal{C}}^L$, we are able to provide an analytic tight bound for the specific codes, as shown in Section III. An easy to track set of necessary conditions for the definition of L-list permutation invariant linear codes is provided next according to the following lemma.

Lemma 2: Let for a non-negative integer $t < N/2 - \lceil d_{\min}/2 \rceil$ and a positive integer s all of the following conditions be valid for a linear code $\mathcal{C}$:

i) every vector $\mathbf{y} \in F_q^N$ of weight $\lceil d_{\min}/2 \rceil + t$ has at least $L+1$ codewords of $\mathcal{C}$ within distance $\lceil d_{\min}/2 \rceil + t$.
ii) every vector $\mathbf{y} \in F_q^N$ of weight $w < \lceil d_{\min}/2 \rceil + t$ has at most L codewords of $\mathcal{C}$ within distance w.
iii) each vector $\mathbf{y} \in F_q^N$ of weight at most $\rho(\mathcal{C})$ has at least $L+1$ codewords of $\mathcal{C}$ within distance $\lceil d_{\min}/2 \rceil + t + s$.

Then $\mathcal{C}$ is L-list permutation invariant.

Proof: If according to Definition 2 we select $w_{opt} = \lceil d_{\min}/2 \rceil + t$, then due to the first statement of the current lemma, according to which every vector $\mathbf{y}$ has at least $L+1$ codewords within distance $\lceil d_{\min}/2 \rceil + t$, property $\mathcal{L}_1$ of Definition 2 will be satisfied. This is due to the fact that the minimum of the elements of the $(\lceil d_{\min}/2 \rceil + t)$-th column of the distance distribution matrix $\mathbf{\Gamma}$ will be at least $L+1$. Moreover, according to the second statement, the sum of all rows of $\mathbf{\Gamma}$ up to the $(\lceil d_{\min}/2 \rceil + t)$-th column, will be at most L. Furthermore, if the third statement of the lemma is satisfied, then for every row (coset weight distribution) of $\mathbf{\Gamma}$ there exists a column s positions right to the $(\lceil d_{\min}/2 \rceil + t)$-th one, such that each vector $\mathbf{y}$ with weight up to the covering radius $\rho(\mathcal{C})$ has at least $L+1$ codewords within distance $\lceil d_{\min}/2 \rceil + t + s$. Consequently, property $\mathcal{L}_2$ of Definition 2 is also satisfied. ∎

Remark 2: For $t = 0$ the second statement of Lemma 2 is trivial, since all vectors $\mathbf{y}$ with weight $w < \lceil d_{\min}/2 \rceil$ have only the zero codeword within distance w. This is the case employed in the analysis next. Specifically, if for $t = 0$ the fixed list size L satisfies

$$L > \mathcal{S}_1, \quad \mathcal{S}_1 \triangleq \sum_{k=d_{\min}}^{\rho(\mathcal{C})+\lceil d_{\min}/2 \rceil + s} S_k \tag{11}$$

where terms S_k are provided by (1), then $\mathcal{C}$ is L-list permutation invariant. Indeed, according to the first statement of Lemma 2, a codeword $\mathbf{c}^*$, apart from the all-zeros one $\mathbf{c}_0$, which is at Hamming distance at most $\lceil d_{\min}/2 \rceil$ from a vector $\mathbf{y}$ of Hamming weight $\mathrm{wt}(\mathbf{y}) = \lceil d_{\min}/2 \rceil$, should satisfy

$$\mathrm{wt}(\mathbf{c}^*) \leq 2\lceil d_{\min}/2 \rceil.$$

The previous statement follows due to the previous assumptions and the Hamming distance inequality according to which

$$\mathrm{wt}(\mathbf{c}^*) - \mathrm{wt}(\mathbf{y}) \leq \lceil d_{\min}/2 \rceil.$$

Therefore, if we set

$$\mathcal{S}_2 \triangleq \sum_{k=d_{\min}}^{2\lceil d_{\min}/2 \rceil} S_k \tag{12}$$

then every vector $\mathbf{y}$ of weight $\lceil d_{\min}/2 \rceil$ enjoys at most $\mathcal{S}_2$ codewords within distance $\lceil d_{\min}/2 \rceil$. With similar arguments, one can deduce that the Hamming weight of each codeword $\mathbf{c}^*$ satisfying the third statement of Lemma 2 is constrained to

$$\mathrm{wt}(\mathbf{c}^*) \leq \rho(\mathcal{C}) + \lceil d_{\min}/2 \rceil + s.$$

Thus the maximum number of codewords a vector $\mathbf{y}$ of weight $\rho(\mathcal{C})$ has within distance $\lceil d_{\min}/2 \rceil + s$ is $\mathcal{S}_1$. Since $\rho(\mathcal{C}) \geq \lceil d_{\min}/2 \rceil$ [18], it holds $\mathcal{S}_2 < \mathcal{S}_1$ and consequently the validity of (11) confirms the L-list permutation invariance property of the linear code $\mathcal{C}$.

A. Existence of L-list permutation invariant linear codes

The construction of general L-list permutation invariant linear codes appears to be a rather difficult process, since it requires the analytic calculation of the distance distribution matrix $\mathbf{\Gamma}$ of the code. In what follows, with the aid of Lemma

2 and for $t = 0$, we investigate the conditions under which such codes appear to exist.

Theorem 1: Let an (N, R) linear code $\mathcal{C}$ satisfy for some positive integer s

$$\min\Big\{\mathcal{S}_1 q^{-N(1-H_q(\delta')-H_q(\rho_r(\mathcal{C})))}, \mathcal{S}_2 (q-1)^{N\delta} \cdot q^{-N(1-2H_q(\delta))}\Big\}(L-1)e < 1 \quad (13)$$

where $\delta = \lceil d_{\min}/2\rceil/N$, $\delta' = (\lceil d_{\min}/2\rceil + s)/N$ and $\mathcal{S}_1$ defined in (11). If $L \in [2, \mathcal{S}_1]$ and $1/2 > \max\{\rho_r(\mathcal{C}), \delta'\}$, then with positive probability $\mathcal{C}$ is L-list permutation invariant.

Proof: Consider an (N, R) random linear code $\mathcal{C}$ over the field F_q constructed by choosing NR linearly independent vectors of length N over F_q. Let $B(\mathbf{y}, r)$ denote the Hamming ball with center $\mathbf{y}$ and radius r, i.e. $B(\mathbf{y}, r) = \{\mathbf{y}' \in F_q^N : \mathrm{wt}(\mathbf{y} - \mathbf{y}') \leq r\}$ with volume $\mathrm{Vol}(B(\mathbf{y}, r))$. If $H_q(p)$ denotes the entropy function $H_q(p) \triangleq -p\log_q p - (1-p)\log_q(1-p)$, then the volume of the considered Hamming ball is upper bounded as

$$\mathrm{Vol}(B(\mathbf{y}, \lceil d_{\min}/2\rceil)) \leq q^{NH_q(\delta)}. \quad (14)$$

Let the following parametrized events, where the all-zeros codeword $\mathbf{c}_0$ is excluded:

$$\mathcal{A}_l = \Big\{\exists\, \mathbf{y} \in F_q^N, \{\mathbf{c}_i^*\}_{i=1}^l \setminus \mathbf{c}_0 : \{\mathbf{c}_i^*\}_{i=1}^l \text{ fall in a Hamming ball } B(\mathbf{y}, \lceil d_{\min}/2\rceil) \text{ with } \mathrm{wt}(\mathbf{y}) = \lceil d_{\min}/2\rceil\Big\}$$

$$\mathcal{B}_l = \Big\{\exists\, \mathbf{y} \in F_q^N, \{\mathbf{c}_i^*\}_{i=1}^l \setminus \mathbf{c}_0 : \{\mathbf{c}_i^*\}_{i=1}^l \text{ fall in a Hamming ball } B(\mathbf{y}, \lceil d_{\min}/2\rceil + s) \text{ where } \mathrm{wt}(\mathbf{y}) \leq \rho(\mathcal{C})\Big\}.$$

We are interested in upper bounding the following series of probabilities $\Pr(\mathcal{A}_l, \mathcal{B}_l)$ for every $l \in [2, L]$, where events $\mathcal{A}_l, \mathcal{B}_l$ are not necessarily disjoint. It is easily deduced that $\Pr(\mathcal{A}_l, \mathcal{B}_l)$ is a decreasing function with respect to l so that we concentrate on upper bounding $\Pr(\mathcal{A}_2, \mathcal{B}_2)$. Due to inequality

$$\Pr(\mathcal{A}_2, \mathcal{B}_2) \leq \min\{\Pr(\mathcal{A}_2), \Pr(\mathcal{B}_2)\} \quad (15)$$

we examine separately the probabilities on the right hand side of the previous inequality. In what follows, we make use of the following argument stated in [19] and in the proof of [20, Theorem 5.6]. Namely, every set of L distinct non-zero messages in F_q^k contains a subset of at least $\ell = \lceil\log_q(L)\rceil$ messages which are linearly independent over F_q. It is easily verified that such linearly independent ℓ-tuples are mapped to ℓ mutually independent codewords by a random linear code.

For the event $\mathcal{A}_2$, the respective probability is upper bounded by the probability there exists a $\mathbf{y} \in F_q^N$ with Hamming weight $\mathrm{wt}(\mathbf{y}) = \lceil d_{\min}/2\rceil$ having $\ell = \lceil\log_q 2\rceil = 1$ codeword $\mathbf{c}_1^*$ out of at most $\mathcal{S}_2$ codewords within Hamming distance $\lceil d_{\min}/2\rceil$, according to Remark 2. The latter probability is upper bounded according to (14) and the union bound as

$$\begin{aligned}\Pr(\mathcal{A}_2) &\leq \mathcal{S}_2 \sum_{\substack{\mathbf{y}\in F_q^N \\ \mathrm{wt}(\mathbf{y}) = \lceil\frac{d_{\min}}{2}\rceil}} \frac{\mathrm{Vol}(B(\mathbf{y}, \lceil d_{\min}/2\rceil)}{q^N} \\ &\leq \mathcal{S}_2 \binom{N}{\lceil\frac{d_{\min}}{2}\rceil}(q-1)^{\lceil\frac{d_{\min}}{2}\rceil}\frac{q^{NH_q(\delta)}}{q^N} \\ &\leq \mathcal{S}_2 (q-1)^{N\delta} q^{-N(1-2H_q(\delta))} \end{aligned} \quad (16)$$

where we have made use of the following relation

$$\binom{n}{k} \leq q^{nH_q\left(\frac{k}{n}\right)}.$$

In a similar manner, the investigated probability of the event $\mathcal{B}_2$ is upper bounded as

$$\begin{aligned}\Pr(\mathcal{B}_2) &\leq \mathcal{S}_1 \sum_{\mathbf{y}:\mathrm{wt}(\mathbf{y})\leq\rho(\mathcal{C})} \binom{N}{\mathrm{wt}(\mathbf{y})}(q-1)^{\mathrm{wt}(\mathbf{y})}\frac{q^{NH_q(\delta')}}{q^N} \\ &\leq \mathcal{S}_1 q^{-N\left(1-H_q(\delta')-H_q(\rho_r(\mathcal{C})\right)} \end{aligned} \quad (17)$$

where the above sum with respect to $\mathrm{wt}(\mathbf{y})$ equals the volume of a Hamming ball with center the all-zero codeword $\mathbf{c}_0$ and radius $\rho(\mathcal{C})$. It is reminded that $\rho_r(\mathcal{C})$ is the relative covering radius of the $\mathcal{C}$. Combing (15)-(17) we have

$$\Pr(\mathcal{A}_2, \mathcal{B}_2) \leq \min\Big\{\mathcal{S}_1 q^{-N(1-H_q(\delta')-H_q(\rho_r(\mathcal{C}))}, \mathcal{S}_2 (q-1)^{N\delta} q^{-N(1-2H_q(\delta))}\Big\}. \quad (18)$$

It is noted that each event $\mathcal{A}_l \cap \mathcal{B}_l$, $l \in [2, L]$ depends on at most $L-1$ others, while the probability of each event is upper bounded by the right hand side of (18). Then according to the symmetric case of Lovasz Local Lemma [13, Corollary 5.1.2], if (13) is satisfied, then with positive probability, none of the events $\mathcal{A}_l \cap \mathcal{B}_l$, $l \in [2, L]$ are valid, and thus the statements of Lemma 2 when $t = 0$ are satisfied. ■

The range of linear codes which satisfy (13) is mainly constrained by the employment of (15) and the use of Lemma 2 for $t = 0$. The latter provide only sufficient conditions for a linear code to be permutation invariant. Moreover, the proof of non-trivial existence results for $t > 0$ requires the development of new bounds in the spirit of [19]. The latter relate the number of codewords of a linear code that fall in a Hamming ball with regard to the ball's radius.

Remark 3: If a linear code $\mathcal{C}$ is 2-list permutation invariant, then it is also ML permutation invariant according to Definition 2. Thus, one can prove the existence of an ML permutation invariant code just by validating (13) for a code $\mathcal{C}$ with $L = 2$.

Remark 4: The positive integer constant s incorporated in (13) through δ' affects the rate of the code $\mathcal{C}$ for which the latter is L-list permutation invariant. Indeed, since $\mathcal{S}_1, \mathcal{S}_2 < M = q^{NR}$, the sufficient condition (13) of Theorem 1 can be replaced by the most stringent inequality

$$\min\Big\{(q-1)^{N\delta} q^{-N(1-2H_q(\delta)-R)}, q^{-N\left(1-H_q(\delta')-H_q(\rho_r(\mathcal{C}))-R\right)}\Big\}(L-1)e < 1. \quad (19)$$

Furthermore, if $\mathcal{C}$ is binary ($q=2$), then (19) reduces to

$$2^{-N(1-2H_2(\delta)-R)}(L-1)e<1 \tag{20}$$

since $1/2 > \min\{\rho_{\rm r}(\mathcal{C}),\delta'\} \geq \delta$ under the assumption of Theorem 1. It is worth mentioning that for large N (20) is approximated for finite L by

$$R<R^* \triangleq 1-2H_2(\delta).$$

Rate R^* is compared next with the capacity of the binary symmetric channel $C=1-H_2(p)$, where p denotes the error transition probability of the channel. One can easily prove by simple differentiation that $H_2(\lambda x)-\lambda H_2(x)$, $\lambda<1/x$, is an increasing function with respect to x when $\lambda<1$ and a decreasing one when $\lambda>1$. If $H_2(p)<H_2(2\delta)$ or $p<2\delta$ then $R^*<C$ since $H_2(2\delta)<2H_2(\delta)$.

Remark 5: For finite list size L and infinite blocklength N, one can still apply Lovasz Local Lemma since the latter applies to a finite number of events $\mathcal{A}_l\cap\mathcal{B}_l$, $l\leq L$. This is not the case when list size $L\to\infty$, and thus the number M of messages is infinite, since one has to employ arguments different from this of compactness[2], to prove the existence of an L-list permutation invariant linear code. Nevertheless, the previous case deserves further investigation since according to the discussion in [19, Subsection II.B] and the references therein, transmission rates very close to the capacity $1-H_q(p)$ of discrete memoryless symmetric channels can be treated, in case L is very large.

III. Error Probability Analysis

Due to the channel symmetry, the average error list decoding probability $P_e^{\mathcal{L}}$ with constant list size L of any linear code $\mathcal{C}$, over the set of messages $\mathcal{M}$ equals $P_{e|0}^{\mathcal{L}}$ [12, Appendix A]. Furthermore, for each code in the constructed permutation ensemble $\mathcal{E}$, message 0 is encoded onto the all zero-codeword $\mathbf{c}_0$ so that $\Pr(\mathbf{y}|\mathbf{c}_0,\mathcal{C})=\Pr(\mathbf{y}|\mathbf{c}_0)$ is valid for all $\mathcal{C}\in\mathcal{E}$. Taking the mean value over the ensemble $\mathcal{E}$ on both sides of (7) and utilizing the error decoding invariance principle of Lemma 1 we have

$$P_e^{\mathcal{L}} \leq \min_{\lambda,\rho\geq 0}\sum_{\mathbf{y}\in\mathbf{Y}_0^{\mathcal{L}}}\Pr(\mathbf{y}|\mathbf{c}_0)E\left[f_{\rm de}(\Omega_L(\mathbf{y},\mathcal{C},\lambda,\rho))\right]. \tag{21}$$

Due to the concavity of the double exponential function $f_{\rm de}(\cdot)$ over $\mathbf{y}\in\mathbf{Y}_0^{\mathcal{L}}$ where $\Omega(\mathbf{y},\mathcal{C},\lambda,\rho)\leq 1$, application of Jensen's inequality leads to

$$P_e^{\mathcal{L}} \leq \min_{\lambda,\rho\geq 0}\sum_{\mathbf{y}\in\mathbf{Y}_0^{\mathcal{L}}}\Pr(\mathbf{y}|\mathbf{c}_0)f_{\rm de}\left(E\left[\Omega_L(\mathbf{y},\mathcal{C},\lambda,\rho)\right]\right). \tag{22}$$

Applying again Jensen's inequality to the convex function $1/x^{\rho}$, $\rho\geq 0$ of $\Omega_L(\mathbf{y},\mathcal{C},\lambda,\rho)$ in (5) gives the following upper bound for $P_e^{\mathcal{L}}$:

$$P_e^{\mathcal{L}} \leq \min_{\lambda,\rho\geq 0}\sum_{\mathbf{y}\in F_q^N}\Pr(\mathbf{y}|\mathbf{c}_0)\cdot \\ f_{\rm de}\left(\frac{L^{\rho}\Pr(\mathbf{y}|\mathbf{c}_0)^{\lambda\rho}}{\left(\sum_{m\neq 0}E\left[\Pr(\mathbf{y}|\mathbf{c}_m,\mathcal{C})^{\lambda}\right]\right)^{\rho}}\right). \tag{23}$$

The mean value appearing in the denominator of the double exponent in (23) is with respect to all possible permutations of the original code $\mathcal{C}$. We extend here the bound of [14] to the case where an (N,R) L-list permutation invariant linear code $\mathcal{C}$ defined on the field F_q is transmitted over a discrete memoryless q-ary output symmetric channel.

Lemma 3: The mean value in the denominator of $f_{\rm de}(x)$ in (23) is upper bounded for all $\rho'\geq 0$ as:

$$\left(E\left[\Pr\left(\mathbf{y}|\mathbf{c}_m,\mathcal{C}\right)^{\lambda}\right]\right)^{\rho} \leq (M-1)^{\rho' P} \\ \left(\sum_{l=0}^{N}\sum_{l_1+\ldots+l_{q-1}=l}b_{l_1,\ldots,l_{q-1}}\left(\frac{v_{l_1,\ldots,l_{q-1}}}{b_{l_1,\ldots,l_{q-1}}}\right)^{Q}\right)^{\frac{\rho' P}{Q}}\cdot \\ \cdot\left(q^{-N}\sum_{\mathbf{x}\in F_q^N}\Pr\left(\mathbf{y}|\mathbf{x}\right)^{\frac{1}{1+\rho'}}\right)^{\rho'} \tag{24}$$

after setting $\lambda=1/(1+\rho')P$ and $\rho=\rho' P$, where

$$b_{l_1,\ldots,l_{q-1}} \triangleq \frac{\binom{N}{l_1,\ldots,l_{q-1}}}{q^N}, \\ v_{l_1,\ldots,l_{q-1}} \triangleq \begin{cases} 0 & l_i=0,\ \forall i\in[1,q-1] \\ \frac{S_{l_1,\ldots,l_{q-1}}}{M-1} & \text{otherwise}\end{cases} \\ P,Q\geq 1,\quad \frac{1}{P}+\frac{1}{Q}=1. \tag{25}$$

The proof of the lemma above is provided in Appendix A. The numerator in $b_{l_1,\ldots,l_{q-1}}$ is the multinomial coefficient. We next apply Lemma 3 to the right hand side of (23) in order to obtain the following analytic expression for the upper bound of an L-list permutation invariant code. The proof of the theorem is given in Appendix B.

Theorem 2: Consider the transmission of an arbitrary set of messages $\mathcal{M}$ over a q-ary symmetric channel with the use of an (N,R) L-list permutation invariant linear code $\mathcal{C}$ with $M=q^{NR}$, defined over the finite field F_q. Then the average L-list error decoding probability, over all messages in $\mathcal{M}$, $P_e^{\mathcal{L}}$ of $\mathcal{C}$ is upper bounded as

$$P_e^{\mathcal{L}} \leq \min_{\rho'\geq 0, P\geq 1}\sum_{l=0}^{N}\binom{N}{l}(1-p)^{N-l}p^l\cdot \\ f_{\rm de}\left(\left(\frac{L}{M-1}\right)^{\rho' P}\frac{\alpha_{p,q}(l,N,\rho')}{\mathcal{F}_{\mathcal{C}}(P)^{\rho'}\beta_{p,q}(N,\rho')}\right) \tag{26}$$

where

$$\alpha_{p,q}(l,N,\rho') \triangleq \left[(1-p)^{N-l}\left(\frac{p}{q-1}\right)^{l}\right]^{\frac{\rho'}{1+\rho'}} \\ \beta_{p,q}(N,\rho') \triangleq \left[\frac{1}{q}(1-p)^{\frac{1}{1+\rho'}}+\frac{q-1}{q}\left(\frac{p}{q-1}\right)^{\frac{1}{1+\rho'}}\right]^{\rho' N} \tag{27}$$

and

$$\mathcal{F}_{\mathcal{C}}(P) \triangleq \\ \left(\sum_{l=0}^{N}\sum_{l_1+\ldots+l_{q-1}=l}b_{l_1,\ldots,l_{q-1}}\left(\frac{v_{l_1,\ldots,l_{q-1}}}{b_{l_1,\ldots,l_{q-1}}}\right)^{\frac{P}{P-1}}\right)^{P-1}. \tag{28}$$

[2]The compactness argument is utilized successfully in the proof of [13, Theorem 5.2.2]

Remark 6: Parameter P in Theorem 2 is noted to control the effect of the distance between the code's weight distribution and this of a fully random block code. Each term of the weight distribution of the latter code equals the multinomial coefficient. For $q = 2$ the bound of Theorem 2 reduces to the bound [14, Theorem 1, eq.(13)].

Remark 7: The analysis provided in the proofs of Lemma 3 and Theorem 2 can be used in order to expand the generalization of the Shulman-Feder bound provided in [2, Appendix A] to the non-binary case.

Remark 8: If the fixed list size L is selected so that $L \leq S_{l_0^*, l_1^*, \ldots, l_{q-1}^*}$,

$$\left\{l_0^*, l_1^*, \ldots, l_{q-1}^*\right\} \triangleq \arg\max_{l_0, \ldots, l_{q-1}} \frac{v_{l_1, \ldots, l_{q-1}}}{b_{l_1, \ldots, l_{q-1}}}. \tag{29}$$

then the optimum value of P in the bound (26) of Theorem 2, as well as in the analysis of [2, Appendix A] which leads to the generalized form of Shulman-Feder bound, is $P_{opt} = 1$. Then, we have

$$\mathcal{F}_{\mathcal{C}} \triangleq \lim_{P \to 1^+} \mathcal{F}_{\mathcal{C}}(P) = \max_{l_0, \ldots, l_{q-1}} \frac{v_{l_1, \ldots, l_{q-1}}}{b_{l_1, \ldots, l_{q-1}}}. \tag{30}$$

In the sequel, we consider the limiting case $P \to 1^+$, since it simplifies the analysis.

Remark 9: Suppose that for very large N, there exists an (N, R) binary linear code $\mathcal{C}$ that is L-list permutation invariant. Then, $\mathcal{C}$ cannot be fully random, i.e. the distance spectrum equals the binomial distribution. Otherwise we could apply inequality $\exp(1 - \exp(x - 1)) \leq 1/x$ to the right hand side of (26) of Theorem 2 for transmission over a binary symmetric channel and get an upper bound over the list error decoding probability that is tighter than the lower bound on the respective error probability provided in [6].

A. Refinement of the bound of Theorem 2

The effectiveness of the bound in (26) for an L-list permutation invariant linear code $\mathcal{C}$ is mainly affected by its weight distribution and the relation of the latter to the multinomial distribution. Indeed, the larger term $\mathcal{F}_{\mathcal{C}}$ in the denominator of the double exponent of (26) is, the looser the specific bound of Theorem 2 becomes for the output symmetric discrete memoryless channels. At least numerically, the proposed bound is noted to be loose for binary linear codes which contain the all '1' codeword (there exists a codeword in which exactly one symbol from $F_q \setminus \{0\}$ is repeated N times). This also appears to be the reason Shulman-Feder bound is loose for some codes.

A linear code $\mathcal{C}$ will not contain a codeword in which exactly one symbol from $F_q \setminus \{0\}$ is repeated N times if and only if there does not exist a row in the parity check matrix of $\mathcal{C}$ which sums up to zero modulo q, where $|F_q| = q$. In case $\mathcal{C}$ is binary, the previous statement is equivalent to the nonexistence of a row in the corresponding parity check matrix with an even number of ones. In what follows, we utilize Gallager's first bounding technique in order to provide a tight bound for the presented class of codes, which enjoy at least one codeword of Hamming weight N. We consider an L-list permutation invariant linear code $\mathcal{C}$ that satisfies the conditions of Lemma 2 for a non-negative integer $t < N/2 - \lceil d_{\min}/2 \rceil$ and a positive constant s such that $s + t < N/2 - \lceil d_{\min}/2 \rceil$. If we set

$$\mathcal{C}_{\mathrm{b}} \triangleq \{\mathbf{c} \in \mathcal{C} : \exists\, k \in F_q \setminus \{0\}, c_i = k\ \forall i \in [1, N]\} \tag{31}$$

the error decoding region $Y_{0,\mathcal{C}}^{\mathcal{L}}$ defined in (3) is equivalently expressed as the union of the following set

$$\begin{aligned} Y_{0,\mathcal{C}_{\mathrm{b}}}^{\mathcal{L}} = \{\mathbf{y} \in F_q^N :\ & \exists \{l_i\}_{i=1}^{L},\ \text{with some } \mathbf{c}_{l_i} \in \mathcal{C}_{\mathrm{b}} \text{ such that} \\ & \Pr(\mathbf{y}|\mathbf{c}_{l_i}, \mathcal{C}) \geq \Pr(\mathbf{y}|\mathbf{c}_0, \mathcal{C}), l_i \neq 0\} \end{aligned} \tag{32}$$

and its complement $\hat{Y}_{0,\mathcal{C}_{\mathrm{b}}}^{\mathcal{L}}$, where every $\mathbf{c} \notin \mathcal{C}_{\mathrm{b}}$. Consider a $\mathbf{y} \in Y_{0,\mathcal{C}_{\mathrm{b}}}^{\mathcal{L}}$ so that there exists a $\mathbf{c} \in \mathcal{C}_{\mathrm{b}}$ with $c_i = k, \forall i \in [1, N]$ satisfying according to (4)

$$\mathrm{wt}(\mathbf{y} - \mathbf{c}) \leq \mathrm{wt}(\mathbf{y}). \tag{33}$$

If we define

$$t_k(\mathbf{y}) \triangleq |\{i : y_i = k\}| \tag{34}$$

then for the considered $\mathbf{y} \in Y_{0,\mathcal{C}_{\mathrm{b}}}^{\mathcal{L}}$ and the specific codeword $\mathbf{c}$ where $c_i = k$ it holds according to (33)

$$\sum_{l=0}^{q-1} t_l(\mathbf{y}) - t_k(\mathbf{y}) \leq \sum_{l=1}^{q-1} t_l(\mathbf{y}) \Leftrightarrow t_0(\mathbf{y}) \leq t_k(\mathbf{y}) \tag{35}$$

where it is noted

$$\sum_{l=0}^{q-1} t_l(\mathbf{y}) = N, \quad \sum_{l=1}^{q-1} t_l(\mathbf{y}) = \mathrm{wt}(\mathbf{y}). \tag{36}$$

In the special case of $F_2 = \{0, 1\}$, it holds

$$t_0(\mathbf{y}) = N - \mathrm{wt}(\mathbf{y}),\ \ t_1(\mathbf{y}) = \mathrm{wt}(\mathbf{y})$$

and (35) gets the form

$$\frac{N}{2} \leq \mathrm{wt}(\mathbf{y}).$$

Under the previous analysis and notations, the subset defined in (32) is equivalently expressed as

$$\begin{aligned} Y_{0,\mathcal{C}_{\mathrm{b}}}^{\mathcal{L}} = \{\mathbf{y} \in F_q^N :\ & \exists \{l_i\}_{i=1}^{L} \Pr(\mathbf{y}|\mathbf{c}_{l_i}, \mathcal{C}) \geq \Pr(\mathbf{y}|\mathbf{c}_0, \mathcal{C}), \\ & l_i \neq 0 \text{ and } t_0(\mathbf{y}) \leq t_k(\mathbf{y}) \text{ for some } k \in F_q \setminus \{0\}\}. \end{aligned} \tag{37}$$

Then the error probability $P_e^{\mathcal{L}}$ is upper bounded as

$$P_e^{\mathcal{L}} \leq \Pr(\mathbf{y} \in \hat{Y}_{0,\mathcal{C}_{\mathrm{b}}}^{\mathcal{L}} | \mathbf{c}_0) + \Pr\left(\bigcup_{k=1}^{q-1} \{t_0(\mathbf{y}) \leq t_k(\mathbf{y})\} \Big| \mathbf{c}_0 \right) \tag{38}$$

where for $\mathbf{y} \in \hat{Y}_{0,\mathcal{C}_{\mathrm{b}}}^{\mathcal{L}}$, $t_0(\mathbf{y}) > t_k(\mathbf{y}), \forall k \in F_q \setminus \{0\}$. Due to the memoryless nature of the treated channels, the second term of the sum in the right hand side of (38) satisfies

$$\begin{aligned} \Pr\left(\bigcup_{k=1}^{q-1} \{t_0(\mathbf{y}) \leq t_k(\mathbf{y})\} \Big| \mathbf{c}_0 \right) &= \sum_{k=1}^{q-1} \Pr\left(t_0(\mathbf{y}) \leq t_k(\mathbf{y}) | \mathbf{c}_0\right) \\ = \sum_{k=1}^{q-1} \sum_{l_0=0}^{N/2} \sum_{\substack{\sum_{i=1}^{q-1} l_i = N - l_0 \\ l_k \geq l_0}} &\binom{N}{l_1, \ldots, l_{q-1}} \prod_{i=0}^{q-1} \Pr(i|0)^{l_i} \end{aligned} \tag{39}$$

where l_i stands for the number of times symbol i appears in sequence $\mathbf{y}$. Under the assumption that the discrete memoryless channels are symmetric, the innermost sum in the right hand side of (39) equals

$$(1-p)^{l_0}\left(\frac{p}{q-1}\right)^{N-l_0}\sum_{\substack{\sum_{i=1}^{q-1} l_i=N-l_0\\ l_k\geq l_0}}\binom{N}{l_1,\ldots,l_k,\ldots,l_{q-1}}.$$

Due to counting arguments it holds for every $k\geq 1$

$$\Pr(t_0(\mathbf{y})\leq t_k(\mathbf{y}))=\Pr(t_0(\mathbf{y})\leq t_1(\mathbf{y})).$$

Thus (39) gets the form

$$\Pr\left(\bigcup_{k=1}^{q-1}\{t_0(\mathbf{y})\leq t_k(\mathbf{y})\}\Big|\mathbf{c}_0\right)=(q-1)\sum_{l_0=0}^{N/2}(1-p)^{l_0}\cdot \left(\frac{p}{q-1}\right)^{N-l_0}\sum_{\substack{\sum_{i=1}^{q-2} l_i=N-l_0\\ l_1\geq l_0}}\binom{N}{l_1,\ldots,l_{q-1}}$$

or equivalently

$$\Pr\left(\bigcup_{k=1}^{q-1}\{t_0(\mathbf{y})\leq t_k(\mathbf{y})\}\Big|\mathbf{c}_0\right)=(q-1)\sum_{l_0=0}^{N/2}\hat{U}_{l_0}(1-p)^{l_0}\cdot \left(\frac{p}{q-1}\right)^{N-l_0},$$ where

$$\hat{U}_{l_0}\triangleq\begin{cases}\binom{N}{l_0} & ,\ q=2\\ \sum_{l_1\geq l_0}\binom{N}{l_0,l_1}(q-2)^{N-l_0-l_1} & ,\ q>2\end{cases}. \quad (40)$$

Regarding the first term of the sum in the right hand side of (38), the event $\mathbf{y}\in\hat{Y}^{\mathcal{L}}_{0,C_\mathrm{b}}$ is equivalent to the statement that none of the codewords, that cause a list decoding error when $\mathbf{y}$ is the received sequence at the output of the channel, belong to the set C_b, defined in (31). Excluding set C_b from the codewords that may cause a list decoding error does not alter the L-list permutation invariance property of the considered code $\mathcal{C}$. Thus one can directly apply Lemma 1 and the bounding technique that leads to Theorem 2 to the first term of the sum in the right hand side of (38).

Indeed, consider a vector $\mathbf{y}$ of Hamming weight $w\in[w_{opt},N/2)$ and let all codewords $\mathbf{c}$ such that $\mathrm{wt}(\mathbf{y}-\mathbf{c})=\mathrm{wt}(\mathbf{y})=w$ (i.e. we examine the particular column w of the code's distance distribution matrix $\mathbf{\Gamma}$). Suppose also there exists a $\mathbf{c}^*\in C_\mathrm{b}$ such that $\mathrm{wt}(\mathbf{y}-\mathbf{c}^*)=\mathrm{wt}(\mathbf{y})$ or equivalently there exists a $k\in F_q\setminus\{0\}$ where according to (34) and (36) it holds

$$t_0(\mathbf{y})=t_k(\mathbf{y}) \text{ with } t_0(\mathbf{y})=N-\mathrm{wt}(\mathbf{y}),$$
$$t_k(\mathbf{y})=\mathrm{wt}(\mathbf{y})-\sum_{\substack{l=1\\ l\neq k}}^{q-1}t_l(\mathbf{y})$$

or equivalently

$$\mathrm{wt}(\mathbf{y})=\frac{N}{2}+\frac{1}{2}\sum_{\substack{l=1\\ l\neq k}}^{q-1}t_l(\mathbf{y}).$$

Under our initial assumption $\mathrm{wt}(\mathbf{y})<N/2$, the relation above and the set definition of $Y^{\mathcal{L}}_{0,C_\mathrm{b}}$ in (37) we deduce that there cannot exist a codeword $\mathbf{c}^*\in C_\mathrm{b}$ such that $\mathbf{y}\in Y^{\mathcal{L}}_{0,C_\mathrm{b}}$. Thus the exclusion of the set C_b from the error decoding region $Y^{\mathcal{L}}_0$ does not alter the first $N/2-1$ columns of the distance distribution matrix $\mathbf{\Gamma}$. Consequently, $\mathcal{C}$ is still an L-list permutation invariant code. Indeed the first two conditions of Lemma 4 are satisfied, and this is also the case with the third one, since we have assumed $s<N/2-\lceil d_{\min}/2\rceil$. In accordance with Theorem 2 the following lemma holds true.

Lemma 4: Under the same assumptions of Theorem 2, the average L-list error decoding probability, over all messages in $\mathcal{M}$, $P^{\mathcal{L}}_e$ of an L-list permutation invariant linear code $\mathcal{C}$, that satisfies conditions of Lemma 2 for a non-negative integer $t<N/2-\lceil d_{\min}/2\rceil$ and a positive constant s such that $s+t<N/2-\lceil d_{\min}/2\rceil$, is upper bounded as

$$P^{\mathcal{L}}_e\leq\min_{\rho'\geq 0}\sum_{l=0}^{N}\binom{N}{l}(1-p)^{N-l}p^l\cdot f_{\mathrm{de}}\left(\left(\frac{L}{M-1}\right)^{\rho'}\frac{\alpha_{p,q}(l,N,\rho')}{\hat{\mathcal{F}}^{\rho'}_{\mathcal{C}}\beta_{p,q}(N,\rho')}\right)+\sum_{l_0=0}^{N/2}\hat{U}_{l_0}(1-p)^{l_0}\left(\frac{p}{q-1}\right)^{N-l_0} \quad (41)$$

where functions $\alpha_{p,q},\beta_{p,q}$ are defined in (27), $\hat{U}_{l_0}$ is defined in (40) and

$$\hat{\mathcal{F}}_{\mathcal{C}}\triangleq\max_{\substack{l_0,\ldots,l_{q-1}\\ l_i<N,\forall i\in[1,q-1]}}\frac{v_{l_1,\ldots,l_{q-1}}}{b_{l_1,\ldots,l_{q-1}}}. \quad (42)$$

Remark 10: The symmetric channel behavior encountered in the bounds of the current section is captured by the generic notions of symmetry in [12] and [21]. In order to apply the double exponential technique to such symmetric channels, code structures different from the Hamming distance distribution matrix of Definition 1 have to be considered, while artificial ensembles, possibly different from the permutation one, must be constructed. On the other hand the new bounds cannot be directly applied to non-symmetric channels. The lack of channel symmetry, as defined in the current work, prevents from treating only the transmission of the all-zeros codeword $\mathbf{c}_0$ in the calculation of the error decoding probability. Furthermore, the permutation invariance property as described in Lemma 1, is not necessarily preserved. Thus for the transmission of each message, we have to consider the largest error decoding region within the class of permuted codes for the specific message, in order to apply the double exponential technique. The latter method is expected to lead to a loss of performance in the deduced error bounds.

IV. ON THE RELIABILITY FUNCTION OF BINARY MEMORYLESS SYMMETRIC CHANNELS

Through the application of inequality $f_{\mathrm{de}}(x)\leq 1/x$ to the right hand side of (26), the list error decoding probability of an L-list permutation invariant linear code $\mathcal{C}$ satisfies for any

$\rho \geq 0$

$$P_e^{\mathcal{L}} \leq \exp\Big(-N \max_{\rho \geq 0} \big\{ E_{\mathrm{o}}(\rho, q) - \rho \left(R - \tfrac{\ln L}{N} + \tfrac{\ln \mathcal{F}_{\mathcal{C}}}{N} \right) \big\} \Big) \quad (43)$$

where

$$E_{\mathrm{o}}(\rho, q) = \rho \ln q - (1+\rho) \ln \left[(q-1) \left(\frac{p}{q-1} \right)^{\frac{1}{1+\rho}} + (1-p)^{\frac{1}{1+\rho}} \right]. \quad (44)$$

The domain of parameter ρ in (43) motivates the quest for specific cases under which the previous bound approaches from above and for very large block length N the sphere packing lower bound [6] on list decoding error probability for discrete memoryless symmetric channels. The latter bound is expressed as

$$P_e^{\mathcal{L}} \geq \exp\Big\{ -N \Big[E_{\mathrm{o}}(\rho^*, q) - \rho^* \left(R - \tfrac{\ln L}{N} - O_1(N) \right) + O_2(N) \Big] \Big\} \quad (45)$$

where ρ^* is the optimum $\rho \geq 0$ that maximizes the exponent in the right hand side of (45) and

$$O_1(N) = \frac{\ln 8}{N} + \frac{q \ln N}{N},\ O_2(N) = \sqrt{\frac{8}{N}} \ln \frac{e}{\sqrt{p_{\min}}} + \frac{\ln 8}{N}.$$

Term $p_{\min}$ expresses the minimum transmission probability of a symbol. For the q-ary symmetric channel it holds $p_{\min} = p/(q-1)$.

Under the scope of (43) and (45) and the restriction that the fixed list size L remains finite according to Remark 5, we are interested in investigating the existence of (N, R) L-list permutation invariant linear codes for which $\mathcal{F}_{\mathcal{C}}$ in (30) is $\ln \mathcal{F}_{\mathcal{C}} = o(N)$, where $o(N)/N$ approaches 0 as N becomes large. In this way, we are able to provide the reliability function of discrete memoryless symmetric channels for these rates under which Theorem 1 is valid. In the current work, we restrict[3] to binary codes. Mimicking the arguments of [22, Section II], we construct a random ensemble of (N, R) binary linear codes by selecting independently, with the same probability each of the N bits of the NR lines of the generator matrix. In what follows, we denote by $D(\xi_1 || \xi_2)$ the Kullback-Leibler distance

$$D(\xi_1 || \xi_2) \triangleq \xi_1 \log_2 \frac{\xi_1}{\xi_2} + (1-\xi_1) \log_2 \frac{1-\xi_1}{1-\xi_2} \quad (46)$$

while $a(N) \doteq b(N)$ is used to denote the asymptotic exponential equivalence of terms $a(N), b(N)$ with respect to N, as this is defined in [22, Section II.A]. The weight distribution $S_{\xi N}$ of an (N, R) typical binary linear block code from the linear code ensemble satisfies for very large N and for every $\epsilon > 0$

$$S_{\xi N} \begin{cases} \doteq 2^{N\left(R - D\left(\xi || \frac{1}{2}\right)\right)}, & \left|\frac{1}{2} - \xi\right| \leq \frac{1}{2} - \xi'(R) - \epsilon \\ = 0, & \left|\frac{1}{2} - \xi\right| \geq \frac{1}{2} - \xi'(R) + \epsilon \end{cases} \quad (47)$$

where $\xi'(R)$ is the minimum distance of the code[4] provided by the solution of

$$D(\xi || \tfrac{1}{2}) - R = 0 \quad (48)$$

with respect to ξ. Nonbinary linear codes with weight distribution characteristics similar to these of (47) also appear in [23, Section 4]. Furthermore, the probability of an (N, R) random linear code $\mathcal{C}$ over F_2 with $\mathcal{F}_N$ being exponential large of order $O(N)$ is upper bounded as

$$\Pr\left(\mathcal{F}_{\mathcal{C}} \geq e^{O(N)} \right) \leq \Pr\left(\exists\, \xi : S_{\xi N} \geq 2^{N\left(R - D\left(\xi || \frac{1}{2}\right) + O(1)\right)} \right). \quad (49)$$

The above relation is easily deduced with the aid of definitions in (25) and (30) respectively, if we employ the following relations for very large N

$$\Pr\left(\mathcal{F}_{\mathcal{C}} \geq e^{O(N)} \right) = \Pr\left(\max_{0 < l \leq N} \frac{S_l}{\binom{N}{l}} \geq 2^{-N}(2^{NR} - 1) e^{O(N)} \right) \leq \Pr\left(\exists\, \xi :\ S_{\xi N} \geq \binom{N}{\xi N} 2^{N(R + O(1) - 1)} \right)$$

and

$$\binom{N}{\xi N} \geq 2^{NH(\xi) - o(N)}, D\left(\xi || \frac{1}{2} \right) = 1 - H_2(\xi).$$

Combining (47) with (49), it is concluded that for very large N the probability in the right hand side of (49) approaches 0 and thus a typical linear code $\mathcal{C}$ from the random linear code ensemble has $\ln \mathcal{F}_{\mathcal{C}} \doteq o(N)$. With the aid of (47) and Theorem 1, we examine the existence of codes from the linear code ensemble that are L-list permutation invariant. In particular, in the asymptotic blocklength regime, parameters $\mathcal{S}_1$ and $\mathcal{S}_2$ defined in (11) and (12) respectively satisfy in accordance with (47) for finite values of parameter $s > 0$,

$$\mathcal{S}_1 \doteq 2^{N\left(R - D\left(\frac{\xi'(R)}{2} + \rho_{\mathrm{r}}(\mathcal{C}) || \frac{1}{2}\right)\right)},\ \ \mathcal{S}_2 = 1. \quad (50)$$

Replacing (50) in Theorem 1, the sufficient condition (13) for the L-list permutation invariance property of a linear code satisfies equivalently

$$\min \left\{ 2^{-N\left[1 - H_2\left(\frac{\xi'(R)}{2}\right) - H_2(\rho_{\mathrm{r}}(\mathcal{C})) + D\left(\frac{\xi'(R)}{2} + \rho_{\mathrm{r}}(\mathcal{C}) || \frac{1}{2}\right) - R\right]}, 2^{-N\left[1 - 2H_2\left(\frac{\xi'(R)}{2}\right)\right]} \right\} < \frac{1}{(L-1)e}. \quad (51)$$

[3]The main reason for not addressing non-binary codes in the current section is the fact that for very large blocklengths N, one can only show code existence results with characteristics similar to these stated in (47). Thus a non-binary L-list permutation invariant linear code does not necessarily possess a normal-like weight distribution for very large N.

[4]In the asymptotic blocklength regime, linear codes $\mathcal{C}$ with rate R and relative minimum distance $\xi'(R) < 1/2$ for which (48) is not valid do not exist.

It is noted that the necessary condition $L < \mathcal{S}_1$ of Theorem 1, which enables the application of Lovasz Local Lemma to the finite case, is satisfied since we consider finite list sizes L, while $\mathcal{S}_1$ is exponentially large according to (50). Consequently, (51) is expressed with the aid of relation $D(x||1/2) = 1 - H_2(x)$ as

$$\max\left\{D\left(\tfrac{\xi'(R)}{2}\big|\big|\tfrac{1}{2}\right) + D\left(\tfrac{\xi'(R)}{2} + \rho_{\rm r}(\mathcal{C})\big|\big|\tfrac{1}{2}\right) - H_2\left(\rho_r(\mathcal{C})\right) - R, 1 - 2H_2\left(\tfrac{\xi'(R)}{2}\right)\right\} > 0. \quad (52)$$

Proposition 1: For all code rates R and relative covering radius $\rho_{\rm r}(\mathcal{C})$ under which (52) is valid, the reliability function of binary memoryless symmetric channels is provided by sphere-packing exponent

$$E(R) \triangleq \sup_{\rho \geq 0}\left(E_{\rm o}(\rho, 2) - \rho R\right)$$

where $E_{\rm o}(\rho, 2)$ is given by (44).
The calculation of the relative covering radius $\rho_r(\mathcal{C})$ of a linear code $\mathcal{C}$ in the asymptotic blocklength regime is difficult and thus we resort to bounds that provide sufficient conditions under which (52) is valid. As an example, we consider the redundancy bound of [24, Proposition 2] for a binary linear code $\mathcal{C}$ according to which $\rho_r(\mathcal{C}) \leq 1 - R$. The Kullback-Leibler distance $D(\xi||\frac{1}{2})$ in (46) as well as $-H(\xi)$ are decreasing with respect to $\xi \leq \frac{1}{2}$. Thus, in case

$$\frac{\xi'(R)}{2} + 1 - R \leq \frac{1}{2} \text{ or } R \geq \frac{1 + \xi'(R)}{2} \quad (53)$$

we have

$$\frac{\xi'(R)}{2} + \rho_r(\mathcal{C}) \leq \frac{1}{2}$$

and consequently (52) holds true for code rates R for which the following function[5]

$$\mathfrak{f}(R) \triangleq \max\left\{D\left(\tfrac{\xi'(R)}{2}\big|\big|\tfrac{1}{2}\right) + D\left(R - \tfrac{\xi'(R)}{2}\big|\big|\tfrac{1}{2}\right) - H_2\left(R\right) - R, 1 - 2H_2\left(\tfrac{\xi'(R)}{2}\right)\right\} \quad (54)$$

is strictly positive. The specific function is illustrated in Fig. 1 for those code rates R which satisfy (53). Namely, for $R \geq 0.5475$, $\mathfrak{f}(R)$ is strictly positive and monotonically increasing. Therefore, for all code rates $R \geq 0.5475$ there exists a respective L-list permutation invariant binary linear code, whose error decoding probability coincides with sphere-packing lower bound. The previous rate region does not necessarily coincide with the critical region [25, eq. (5.6.30)] for which the reliability function of the binary symmetric channel is known. For example, let the binary symmetric channel with error transition probability $p = 0.005$. Then, the reliability function of the channel equals $\sup_{\rho\in[0,1]}\left(E_o(\rho, 2) - \rho R\right)$ for all rates $R \in [R_c(5\times10^{-3}), 1 - H_2(5\times10^{-3})]$, where $R_c(p)$ is the critical rate of the channel given as

$$2R_c(p) \triangleq 1 - \log_2\left(\sqrt{1-p} + \sqrt{p}\right) + \frac{\sqrt{1-p}\log_2(1-p) + \sqrt{p}\log_2(p)}{2\left(\sqrt{1-p} + \sqrt{p}\right)}. \quad (55)$$

[5]In the binary case, $D(x||1/2)$ and $H_2(x)$ are symmetric around $1/2$. Thus $D(1-x||1/2) = D(x||1/2)$ and $H_2(1-x) = H_2(x)$.

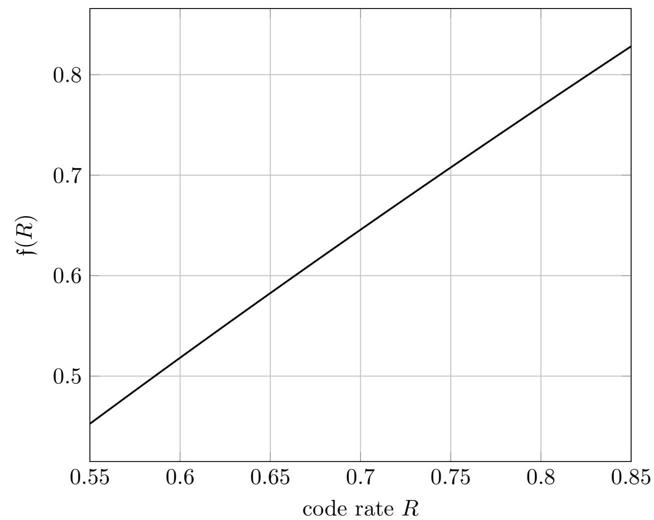


Fig. 1. Examination of the positiveness of $\mathfrak{f}(R)$ in (54) for code rates R which satisfy (53).

Proposition 1 extends the previous region to $[0.5475, 1 - H_2(5\times10^{-3})]$ since $R_c(5\times10^{-3}) = 0.648437$.

The proof of existence of L-list permutation invariant linear codes of finite blocklength N with error decoding probability close to improved sphere packing bounds of [7] and [8] deserves further investigation as it requires an analytic construction process. In general, the proposed bound (43) is able to approach the sphere-packing bound (45) since the optimization parameter ρ, that controls the magnitude of the sum of likelihood ratios $1/\Omega(\mathbf{y}, \mathcal{C}, \lambda, \rho)$ for a vector $\mathbf{y}$, is not constrained to $[0, 1]$. To achieve this, special structure is imposed on the examined codes as well as the artificially constructed ensembles, so that Lemma 1 becomes valid and parameter $\ln \mathcal{F}_{\mathcal{C}}$ in (28) is of order $o(N)$. One can argue that the double exponential function substitutes ρ in mitigating the effect of large values in $E[1/\Omega(\mathbf{y}, \mathcal{C}, \lambda, \rho)]$ and thus ρ can be greater than 1.

V. Comparisons with previous bounds: A new Class of Bounding Techniques

The bounding technique which is presented in Section II and leads to bound (7) can be applied to general linear codes with no particular restrictions, such as list permutation invariance. The resulting bound appears to be tighter than the DS2 bound for list decoding, but its computability comes at the extra cost of requiring in-depth knowledge of the code's characteristics. Indeed, if we extend the sum in (7) over the set F_q^N, the average list error decoding probability of the linear code $\mathbf{C}$ over the set of messages $\mathcal{M}$ is upper bounded as

$$P_{e|0}^{\mathcal{L}} \leq \min_{\lambda,\rho\geq 0} \sum_{\mathbf{y}\in F_q^N} \Pr(\mathbf{y}|\mathbf{c}_0)\cdot f_{\rm de}\left(\frac{L^\rho \Pr(\mathbf{y}|\mathbf{c}_0)^{\lambda\rho}}{\left(\sum_{m\neq 0}\Pr(\mathbf{y}|\mathbf{c}_m)^\lambda\right)^\rho}\right). \quad (56)$$

$$P_{e|0}^{L} \leq \min_{\substack{\lambda\geq 0,\rho\in[0,1]\\ g_0(y):\{0,1\}\to\mathbb{R}}} \frac{1}{L^{\rho}} \left((1-p)g_0(0)+pg_0(1)\right)^{N(1-\rho)} \left(\sum_{d=d_{\min}}^{N} S_d \left((1-p)g_0(0)^{1-\frac{1}{\rho}}+pg_0(1)^{1-\frac{1}{\rho}}\right)^{N-d} \left(\sum_{r=0}^{d}\binom{d}{r}\left((1-p)g_0(0)^{1-\frac{1}{\rho}}\right)^{d-r}\left(pg_0(1)^{1-\frac{1}{\rho}}\right)^{r} f_{\text{de}}\left(\left(\frac{p}{1-p}\right)^{\lambda(2r-d)}\right)\right)\right)^{\rho}. \quad (63)$$

Since the treated channels are memoryless and symmetric, we have in line to the proof of Theorem 2

$$\frac{\Pr(\mathbf{y}|\mathbf{c}_0)^{\lambda\rho}}{\left(\sum_{m\neq 0}\Pr(\mathbf{y}|\mathbf{c}_m)^{\lambda}\right)^{\rho}} = \left(\sum_{m\neq 0}\left(\frac{p}{(q-1)(1-p)}\right)^{\lambda(\text{wt}(\mathbf{y}-\mathbf{c}_m)-\text{wt}(\mathbf{y}))}\right)^{-\rho}. \quad (57)$$

Furthermore, for a fixed $\mathbf{y}\in F_q^N$, sequence $\{\text{wt}(\mathbf{y}-\mathbf{c}_m)\}_{m=0}^{M-1}$ equals the weight distribution of the coset that $\mathbf{y}$ belongs to, and is given by the $i(\mathbf{y})$-th row of matrix $\mathbf{\Gamma}$ according to Definition 1. Thus the right hand side of (57) equals

$$\left(\sum_{l=0}^{N}\mathbf{\Gamma}_{i(\mathbf{y}),l}\left(\frac{p}{(q-1)(1-p)}\right)^{\lambda(l-\text{wt}(\mathbf{y}))}\right)^{-\rho}. \quad (58)$$

If for each row i of the coset weight distribution matrix $\mathbf{\Gamma}$ we group together those vectors $\mathbf{y}\in F_q^N$ of the same weight w and denote their number by $\mathbf{\Gamma}_{i,l}^{(w)}$, the upper bound of (56) on $P_{e|0}^{\mathcal{L}}$ is expressed due to (57)-(58) and the reasoning that leads to (88) in Appendix B as

$$P_{e|0}^{\mathcal{L}} \leq \min_{\lambda,\rho\geq 0}\sum_{i=1}^{K}\sum_{w=0}^{N}\mathbf{\Gamma}_{i,l}^{(w)}(1-p)^{N-w}\left(\frac{p}{q-1}\right)^{w}\cdot f_{\text{de}}\left(\left(\frac{1}{L}\sum_{l=0}^{N}\mathbf{\Gamma}_{i,l}\left(\frac{p}{(q-1)(1-p)}\right)^{\lambda(l-w)}\right)^{-\rho}\right). \quad (59)$$

The DS2 bounding technique for discrete memoryless channels can be applied to the right hand side of (56) and thus lead to a bound similar to (59), which is tighter than the classical DS2 bound for list decoding. Nevertheless, its calculation requires details (sequences $\mathbf{\Gamma}_{i,l}$ and $\mathbf{\Gamma}_{i,l}^{(w)}$) about the structure of the linear code $\mathcal{C}$, which are finer than these required for the exact calculation of error decoding probability. The above discussion identifies the difficulties in the exploration of new efficient bounding techniques, which provide bounds tighter than the ones mentioned in the literature [1].

Identifying the order of summation in (56) with respect to $\mathbf{y}\in F_q^N$ and $m\in[1,M-1]$ as the main reason that leads to the extremely complicated bound (59), we are able to circumvent the previous difficulties and provide a new tight bound. Thus, instead of following the analysis of (3)-(6), we note from (3) that for $\lambda,\rho\geq 0$

$$\text{if } \mathbf{y}\in\mathbf{Y}_{0,\mathcal{C}}^{L},\ 1\leq\left(\frac{1}{L}\sum_{m\neq 0}f_{\text{de}}\left(\left(\frac{\Pr(\mathbf{y}|\mathbf{c}_0)}{\Pr(\mathbf{y}|\mathbf{c}_m)}\right)^{\lambda}\right)\right)^{\rho}. \quad (60)$$

If we combine (60) with (2), we get the following bound on the error decoding probability for the specific linear code $\mathcal{C}$

$$P_{e|0}^{L} \leq \min_{\lambda,\rho\geq 0}\sum_{\mathbf{y}\in F_q^N}\Pr(\mathbf{y}|\mathbf{c}_0)\cdot\left(\frac{1}{L}\sum_{m\neq 0}f_{\text{de}}\left(\left(\frac{\Pr(\mathbf{y}|\mathbf{c}_0)}{\Pr(\mathbf{y}|\mathbf{c}_m)}\right)^{\lambda}\right)\right)^{\rho} \quad (61)$$

where the sum with respect to $\mathbf{y}$ is expanded from $\mathbf{Y}_{0,\mathcal{C}}^{L}$ to F_q^N. It is noted here that a direct comparison between (61) and the bound of (7) with the sum extended over F_q^N is not straightforward. Applying the DS2 bounding technique [2, Sec. II.B] to the right hand side of (61) under the assumption $\rho\leq 1$, we get for a non-negative un-normalized tilting measure $G_0(\mathbf{y})$

$$P_{e|0}^{L} \leq \min_{\substack{\lambda\geq 0\\ \rho\in[0,1]}}\frac{1}{L^{\rho}}\left(\sum_{\mathbf{y}\in F_q^N}G_0(\mathbf{y})\Pr(\mathbf{y}|\mathbf{c}_0)\right)^{1-\rho}\left(\sum_{m\neq 0}\sum_{\mathbf{y}\in F_q^N}\Pr(\mathbf{y}|\mathbf{c}_0)G_0(\mathbf{y})^{1-\frac{1}{\rho}}f_{\text{de}}\left(\left(\frac{\Pr(\mathbf{y}|\mathbf{c}_0)}{\Pr(\mathbf{y}|\mathbf{c}_m)}\right)^{\lambda}\right)\right)^{\rho}. \quad (62)$$

Due to inequality $f_{\text{de}}(x)\leq 1/x$, it follows that the right hand side of (62) is tighter than the classical DS2 bound and all its modifications related to Shulman-Feder bound, as reported in [26] and [12]. In what follows we provide an expression for the tight bound of (62) in the case of binary memoryless output symmetric channels with $F_2=\{0,1\}$, and $G_0(\mathbf{y})=\prod_{i=1}^{N}g_0(y_i)$ for a positive function $g_0(y):\{0,1\}\to\mathbb{R}$. The general alphabet case is more complicated and is provided in Appendix C. The analysis for binary input ternary output has recently appeared in [27].

Theorem 3: Consider the transmission of a set of messages $\mathcal{M}$ with cardinality M with the use of an (N,R) binary linear block code $\mathcal{C}$ over a binary symmetric channel with error transition probability p. If the weight distribution of $\mathcal{C}$ is provided by the sequence $\{S_d\}_{d=d_{\min}}^{N}$ where $d_{\min}$ is the minimum distance of the code, then the average list decoding error probability over $\mathcal{M}$ is upper bounded as shown by (63) at the top of the current page.

Proof: The base of the first term of the product in the right hand side of (62), which is raised to the $1-\rho$ power, depends only on the transmitted codeword $\mathbf{c}_0$ and equals according to the assumptions of the theorem

$$\sum_{\mathbf{y}\in F_q^N}G_0(\mathbf{y})\Pr(\mathbf{y}|\mathbf{c}_0)=\sum_{i=1}^{N}\sum_{y_i\in F_q}\prod_{i=1}^{N}g_0(y_i)\Pr(y_i|0)=$$

$$\prod_{i=1}^{N}\sum_{y_i\in F_q} g_0(y_i)\Pr(y_i|0) = \left(\sum_{y\in F_q} g_0(y)\Pr(y|0)\right)^N$$
$$= \left((1-p)g_0(0)+pg_0(1)\right)^N . \tag{64}$$

For the second term of the product in the right hand side of (62), we treat each term of the sum with respect to m separately. Specifically, consider a specific codeword $\mathbf{c}_m$ of $\mathcal{C}$ with Hamming weight d and let the subsets I_0^m and I_1^m consist of the positions in the sequence $\mathbf{c}_m$ where symbols 0 and 1 appear respectively, so that $|I_0^m| = N-d$ and $|I_1^m| = d$. Then, the double exponent in the right hand side of (62) satisfies

$$\left(\frac{\Pr(\mathbf{y}|\mathbf{c}_0)}{\Pr(\mathbf{y}|\mathbf{c}_m)}\right)^\lambda = \prod_{i\in I_1^m}\left(\frac{\Pr(y_i|0)}{\Pr(y_i|1)}\right)^\lambda \tag{65}$$

and is independent of the positions $i \in I_0^m$. Thus, we have

$$\sum_{\mathbf{y}\in F_q^N}\Pr(\mathbf{y}|\mathbf{c}_0)G_0(\mathbf{y})^{1-\frac{1}{\rho}} f_{\text{de}}\left(\left(\frac{\Pr(\mathbf{y}|\mathbf{c}_0)}{\Pr(\mathbf{y}|\mathbf{c}_m)}\right)^\lambda\right) =$$
$$\left(\sum_{i\in I_0^m}\sum_{y_i\in F_q}\prod_{i\in I_0^m}\Pr(y_i|0)g_0(y_i)^{1-\frac{1}{\rho}}\right)\left(\sum_{i\in I_1^m}\sum_{y_i\in F_q}\prod_{i\in I_1^m}\right.$$
$$\left.\Pr(y_i|0)g_0(y_i)^{1-\frac{1}{\rho}} f_{\text{de}}\left(\prod_{i\in I_1^m}\left(\frac{\Pr(y_i|0)}{\Pr(y_i|1)}\right)^\lambda\right)\right). \tag{66}$$

The first product in the right hand side of (66) satisfies

$$\left(\sum_{y\in F_q}\Pr(y|0)g_0(y)^{1-\frac{1}{\rho}}\right)^{N-d} =$$
$$\left((1-p)g_0(0)^{1-\frac{1}{\rho}}+pg_0(1)^{1-\frac{1}{\rho}}\right)^{N-d}. \tag{67}$$

For the second term of the product in the right hand side of (66) consider a specific value of the subsequence $\{y_i\}_{i\in I_1^m}$ which consists of r symbols 1 and $d-r$ symbols 0. For this value, the respective summand in the examined term of the product amounts to

$$\left(\Pr(0|0)g_0(0)^{1-\frac{1}{\rho}}\right)^{d-r}\left(\Pr(1|0)g_0(1)^{1-\frac{1}{\rho}}\right)^r\cdot$$
$$f_{\text{de}}\left(\left(\frac{\Pr(0|0)}{\Pr(0|1)}\right)^{\lambda(d-r)}\left(\frac{\Pr(1|0)}{\Pr(1|1)}\right)^{\lambda r}\right) = \left(pg_0(0)^{1-\frac{1}{\rho}}\right)^r\cdot$$
$$\left((1-p)g_0(0)^{1-\frac{1}{\rho}}\right)^{d-r} f_{\text{de}}\left(\left(\frac{p}{q-1}\right)^{2\lambda r-\lambda d}\right). \tag{68}$$

Since the right hand side of (68) is noted to depend only the Hamming weight r of the subsequence $\{y_i\}_{i\in I_1^m}$, $\binom{d}{r}$ subsequences contribute in the manner of (68) in the sum of the second term of the product in (66) and thus the latter sum equals

$$\sum_{r=0}^{d}\binom{d}{r}\left((1-p)g_0(0)^{1-\frac{1}{\rho}}\right)^{d-r}\left(pg_0(0)^{1-\frac{1}{\rho}}\right)^r$$
$$f_{\text{de}}\left(\left(\frac{p}{q-1}\right)^{2\lambda r-\lambda d}\right). \tag{69}$$

The proof of the theorem is now complete if in (62) we substitute (64) as well as the result of the combination of (69) with (67) along with the note that the product of these two terms is related to the codeword $\mathbf{c}_m$ only through it's Hamming weight d. ■

Remark 11: If we apply inequality $f_{\text{de}}(x) \le 1/x$ to the right hand side of (63), then one gets the DS2 bound for the binary symmetric channel. Indeed, if we set

$$\gamma_1^{\mathcal{C}}(\rho) \triangleq \frac{1}{L^\rho}\left((1-p)g_0(0)+pg_0(1)\right)^{N(1-\rho)}$$
$$\gamma_2^{\mathcal{C}}(\rho,d) \triangleq \left((1-p)g_0(0)^{1-\frac{1}{\rho}}+pg_0(1)^{1-\frac{1}{\rho}}\right)^{N-d}$$

then we have

$$P_{e|0}^L \le \gamma_1^{\mathcal{C}}(\rho)\left(\sum_{d=d_{\min}}^{N} S_d\gamma_2^{\mathcal{C}}(\rho,d)p^{\lambda d}(1-p)^{\lambda d}\left(\sum_{r=0}^{d}\binom{d}{r}\cdot\right.\right.$$
$$\left.\left.\left((1-p)^{1-2\lambda}g_0(0)^{1-\frac{1}{\rho}}\right)^{d-r}\left(p^{1-2\lambda}g_0(1)^{1-\frac{1}{\rho}}\right)^r\right)\right)^\rho =$$
$$\gamma_1(\rho)\left(\sum_{d=d_{\min}}^{N} S_d\gamma_2^{\mathcal{C}}(\rho,d)p^{\lambda d}(1-p)^{\lambda d}\cdot\right.$$
$$\left.\left((1-p)^{1-2\lambda}g_0(0)^{1-\frac{1}{\rho}}+p^{1-2\lambda}g_0(1)^{1-\frac{1}{\rho}}\right)^d\right)^\rho =$$
$$\gamma_1^{\mathcal{C}}(\rho)\left(\sum_{d=d_{\min}}^{N} S_d\gamma_2^{\mathcal{C}}(\rho,d)\cdot\right.$$
$$\left.\left(p^\lambda(1-p)^{1-\lambda}g_0(0)^{1-\frac{1}{\rho}}+p^{1-\lambda}(1-p)^\lambda g_0(1)^{1-\frac{1}{\rho}}\right)^d\right)^\rho .$$

Remark 12: Due to concavity of x^ρ, $\rho\in[0,1]$, the bound of Theorem 3 can be applied to random ensembles of codes, where the average weight distribution is easier to calculate than in the stand-alone case.

Remark 13: Comparing the double exponential function $f_{\text{de}}(x)$ with $f_1(x) \triangleq 1/x$ for $x>0$, as illustrated in Fig. 2, $f_{\text{de}}(x)$ is noted to mitigate the effect of codewords $\mathbf{c}_m$ in the sum (60) for which the likelihood ratio $\Pr(\mathbf{y}|\mathbf{c}_0)/\Pr(\mathbf{y}|\mathbf{c}_m) < 1$ ($\text{wt}(\mathbf{y}-\mathbf{c})-\text{wt}(\mathbf{y})<0$ respectively) for a given $\mathbf{y}\in F_q^N$. The question whether a horizontally flip sigmoid function different from the double exponential function and strictly lower than or equal to $f_1(x)$ can be employed in the DS2 bounding technique naturally arises. Sigmoid functions interpreted as cumulative distribution functions of continuous probability functions appear to be good candidates. In the sequel we consider the normalized horizontally flip sigmoid

$$f_2(x) \triangleq \frac{\text{erfc}\left(0.5e^{x-1}\right)}{\text{erfc}(0.5)} \tag{70}$$

depicted in Fig. 2, where

$$\text{erfc}(x) = 1-\sqrt{\frac{2}{\pi}}\int_0^{\sqrt{2}x} e^{-\frac{z^2}{2}}dz.$$

The scaling factor 0.5 is selected so that $f_2(x)$ is as close as possible to $f_1(x)$ in the region close to 1. We refer to the upper bound (63) where $f_{\text{de}}(x)$ is replaced by $f_2(x)$ as (63)-erfc. It is noted that $f_2(x)$ is slightly larger than $f_1(x)$ in the region

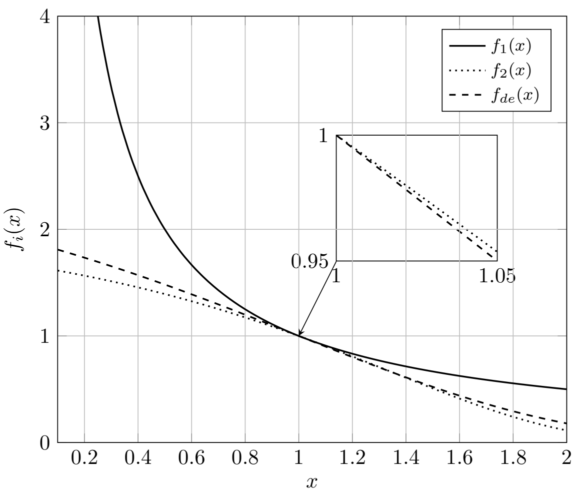


Fig. 2. Horizontally flip sigmoid functions $f_{\text{de}}(x)$ and $f_2(x) \triangleq \operatorname{erfc}(0.5\exp(x-1))/\operatorname{erfc}(0.5)$ used to replace $f_1(x) \triangleq x^{-1}, x > 0$ in the classical DS2 bound.

around 1 and lower than $f_1(x)$ and $f_{\text{de}}(x)$ in the complement of the previous region.

Remark 14: Any horizontally flip sigmoid function $s(x)$ with behavior similar to those of Fig. 2 constitutes an upper bound to the following version of the unit step function $u(x) = 1, x \in [0,1]$, $u(x) = 0, x > 1$. Even though $u(x)$ appears as the tightest $s(x)$ one can employ to replace $f_1(x)$ in the DS2 bounding technique, nevertheless it fails to provide analytic bound expressions. Furthermore, any smooth $s(x)$ with $(1-s(x))(1-x) \leq 0, x \geq 0$ and $s(x) \leq 1/x$ provably approximates the probability of the error decoding region in a manner more efficient than $f_1(x)$.

According to Remarks 12 and 14, we are able to apply the DS2 bounding technique to the random ensemble of codes [5] as in [2], with $f_1(x)$ replaced by any $s(x)$ that satisfies the properties of the previous remark.

Proposition 2: The average probability of error $\overline{P}_e$ over the ensemble of random codes and over all messages is upper bounded for every function $s(x) : [0,\infty) \to [0,\infty)$, $(1-s(x))(1-x) \leq 0$ as

$$\overline{P}_e \leq \min_{\substack{\lambda \geq 0 \\ \rho \in [0,1]}} M^\rho \sum_{\mathbf{y}} \left(\sum_{\mathbf{x}} q_N(\mathbf{x}) \Pr(\mathbf{y}|\mathbf{x}) \right) \cdot \left(\sum_{\mathbf{x}'} q_N(\mathbf{x}') s\left(\frac{\Pr(\mathbf{y}|\mathbf{x})^\lambda}{\Pr(\mathbf{y}|\mathbf{x}')^\lambda} \right) \right)^\rho \quad (71)$$

where $q_N(x_m)$ denotes the probability distribution with which we map a message m to a codeword $\mathbf{x}_m$ in the code ensemble. The un-normalized tilting distribution $G_N^m(\mathbf{y})$ that minimizes the random coding version of the DS2 bound, in analogy to [2, eq.(10)], and leads to (71) is given by

$$G_N^m(\mathbf{y}) = \left(\sum_{\mathbf{x}'} q_N(\mathbf{x}') s\left(\frac{\Pr(\mathbf{y}|\mathbf{x}_m)^\lambda}{\Pr(\mathbf{y}|\mathbf{x}')^\lambda} \right) \right)^\rho . \quad (72)$$

Gallager's random coding theorem results now as a special case of (71) if we select $s(x) = 1/x$ and set $\lambda = 1/(1+\rho)$. The quest for proper forms of functions $s(x)$ which are able to provide tight bounds in the finite blocklength regime (for the average code the random coding bound is tight [28]) requires further analysis.

Remark 15: The previous analysis can be applied in order to derive upper bounds on the bit error decoding probability of q-ary linear systematic codes. In particular, for a binary linear systematic code $\mathcal{C}$, as stated in [1, Section 4.8], the upper bound (63) of Theorem 3 can be used as an upper bound to the bit error decoding probability of $\mathcal{C}$ with the weight distribution sequence $\{S_d\}_{d=d_{\min}}^N$ replaced by the input-output weight enumeration sequence $\{S'_d\}_{d=d_{\min}}^N$ where

$$S'_d = \sum_{w=1}^{NR} \frac{w}{NR} S'_{w,d}$$

and $S'_{w,d}$ designates the number of codewords of $\mathcal{C}$ with information weight w and Hamming weight d.

VI. Applications

In the current section, we consider the transmission of the binary Reed-Muller RM$(2,5)$ code over a discrete memoryless binary symmetric channel with error transition probability p as well as the transmission of the ternary $[13,7,5]$ quadratic residue (QR) [17] over a discrete memoryless 3-ary symmetric channel, with error transition probability $p/2$. At the output of the considered channels, we perform list decoding under fixed list sizes L_1 and L_2 respectively. With the aid of the distance distribution matrices illustrated in [18, Table 11.4] and Table I, the latter list size values are constrained to the following sets $L_1 \in \mathfrak{L}_1 = \{5,16\}$ and $L_2 \in \mathfrak{L}_2 = \{1,6\}$ so that the considered RM and QR codes are L_1-list and L_2-list permutation invariant respectively. Under the previous communication setups, we calculate the upper bounds on list decoding error probability of Theorem 2, Lemma 4 and Theorem 3. For ease of analysis we allow $P \to 1$ and utilize accordingly the terms $\mathcal{F}_\mathcal{C}$ and $\hat{\mathcal{F}}_\mathcal{C}$ defined in (28) and (42) respectively. We next compare the previous bounds with the list decoding version of the DS2 bound as well as the generalized version of Shulman-Feder (GSF) bound. The latter gets the following form [2, eq. (A17)] for $P \to 1$

$$P_e \leq \min_{\rho \in [0,1]} \frac{\left((M-1)\hat{\mathcal{F}}_\mathcal{C}\right)^\rho}{q^{\rho N}} \cdot \left(\sum_y \left(\sum_x \Pr(y|x)^{\frac{1}{1+\rho}} \right)^{1+\rho} \right)^N . \quad (73)$$

It is noted that both codes satisfy the conditions of Lemma 4 regarding the treated list sizes. Indeed, the RM$(2,5)$ code, in case of $L_1 = 5$ satisfies Lemma 4 for $(t,s) = (0,4)$ and the same is valid in case of $L_1 = 16$ for $(t,s) = (2,2)$. In addition, the QR code for $L_2 = 1$ satisfies Lemma 4 for $(t,s) = (0,1)$, while the same holds true for $L_2 = 5$ under $(t,s) = (1,1)$. The numerical optimization of all bounds was performed[6] with the aid of "SciPy" package in Python.

[6]Due to the nonlinear behavior of the examined bounds with respect to the parameters under optimization, the minimization process is repeated with different initial values and the minimum is taken.

Initially we consider the case of the $\mathrm{RM}(2,5)$ code and evaluate numerically the bounds (26), (41) and (63) along with the DS2 bound presented in the equation of Remark 11 and the GSF bound of (73) for fixed list sizes $L_1 \in \mathfrak{L}_1$. Results are depicted in Figs. 3a, 3b for the two cases of fixed lists size $L_1 = 5$ and $L_1 = 16$ respectively in relation with error transition p of the binary symmetric channel. For $L_1 = 5$, bound (63) of Theorem 3 is the tightest one, while the bound (41) of Lemma 4 is close to the optimized version of the DS2 bound with respect to parameters λ, ρ, g_0, g_1. For $L_1 = 16$, there exists a region of error tradition probabilities where the bound of Lemma 4 is the tightest one. For both list sizes, the bound of Theorem 2 is noted to be loose. This is due to the fact that $\mathcal{F}_{\mathcal{C}} \simeq 65537$. This appears to be the case also for the QR code where $\hat{\mathcal{F}}_{\mathcal{C}} \simeq 729$.

Next we treat the case where the $[13,7,5]$ Quadratic Residue QR code is transmitted over a 3-ary symmetric channel with error transition probability $p/2$. For the general q-ary case, the average list decoding error probability is proved in Appendix C to satisfy

$$P_{e|0}^{L} \leq \min_{\substack{\lambda \geq 0 \\ \rho \in [0,1]}} \frac{1}{L^{\rho}} \left(\sum_{y \in F_q} \Pr(y|0) g_0(y) \right)^{N(1-\rho)} \cdot \left(\sum_{l=d_{\min}}^{N} \left(\sum_{y \in F_q} \Pr(y|0) g_0(y)^{1-\frac{1}{\rho}} \right)^{N-l} \cdot \sum_{l_1+\ldots+l_{q-1}=l} S_{l_1,\ldots,l_q} f_3\left(\lambda, \rho, \{l_i\}_{i=1}^{q-1}\right) \right)^{\rho}. \quad (74)$$

Function f_3 is defined in (106) and satisfies for the specific channel,

$$f_3(\lambda,\rho,\{l_i\}_{i=1}^2) = \sum\sum \binom{l_1}{|Y_{0,I_1^m}|, |Y_{1,I_1^m}|} \cdot \binom{l_2}{|Y_{0,I_2^m}|, |Y_{2,I_2^m}|} f_1(\lambda,\rho,\{|Y_{0,I_j^m}|\}_{j=1}^2, \{|Y_{j,I_j^m}|\}_{j=1}^2) \cdot f_2(\rho,\{l_i\}_{i=1}^2, \{|Y_{j,I_j^m}|\}_{j=1}^2, \{|Y_{0,I_j^m}|\}_{j=1}^2). \quad (75)$$

The first sum is with respect to all possible values of $|Y_{0,I_1^m}|, |Y_{1,I_1^m}|$ and $|Y_{2,I_1^m}|$ that sum up to l_1 while the second sum is with respect to all $|Y_{0,I_2^m}|, |Y_{1,I_2^m}|$ and $|Y_{2,I_2^m}|$ that sum up to l_2 respectively. Furthermore, functions f_1, f_2, defined in (103) and (105) respectively, satisfy for the specific channel

$$f_1(\lambda,\rho,\{|Y_{j,I_j^m}|\}_{i=1}^2, \{|Y_{0,I_j^m}|\}_{i=1}^2) = \left(\frac{p}{2} g_0(1)\right)^{|Y_{1,I_1^m}|} \cdot \left(\frac{p}{2} g_0(2)\right)^{|Y_{2,I_2^m}|} ((1-p) g_0(0)^{1-\frac{1}{\rho}})^{|Y_{0,I_1^m}|+|Y_{0,I_2^m}|} \cdot f_{\mathrm{de}}\left(\left(\frac{p}{2(1-p)}\right)^{\lambda\left(|Y_{1,I_1^m}|-|Y_{0,I_1^m}|+|Y_{2,I_2^m}|-|Y_{0,I_2^m}|\right)} \right)$$

and

$$f_2(\rho,\{l_i\}_{i=1}^2, \{|Y_{j,I_j^m}|\}_{i=1}^2, \{|Y_{0,I_j^m}|\}_{i=1}^2) = \left(\frac{p}{2} g_0(1)^{1-\frac{1}{\rho}}\right)^{X_1} \left(\frac{p}{2} g_0(2)^{1-\frac{1}{\rho}}\right)^{X_2}$$

where $X_1 = l_1 - |Y_{1,I_1^m}| - |Y_{0,I_1^m}|$ and $X_2 = l_2 - |Y_{2,I_2^m}| - |Y_{0,I_2^m}|$.

Numerical results of the optimized bounds with respect to the free parameters are presented in Figs. 4a and 4b for the permutation invariant list size cases $L_2 = 1$ and $L_2 = 6$ respectively. In both figures, the generalized SF bound is not depicted since it is very close to 1. This is also the case with the bound (26) of Theorem 2 when $L_2 = 1$. Numerical imprecisions during the numerical minimization process are responsible for the non-monotonic behavior of the curve of Theorem 3 in Fig. 4b as well as the curly curve of the optimized DS2 bound in Fig. 4a.

Finally, we study the difference between the bound in (63) and the bound in (63) when the double exponential is replaced with the horizontally flip sigmoid function in (70). We consider again the transmission of the $\mathrm{RM}(2,5)$ over a binary symmetric channel with error transition probability p under list decoding with fixed list size $L = 1$. The difference is illustrated in Fig. 5 to be extremely small. Thus there appears to be no loss in the analysis if we employ a horizontally flip sigmoid function $s(x)$, different from the double exponential function $f_{\mathrm{de}}(x)$, which is not necessarily strictly lower than $f_1(x) = 1/x$. The previous observation can appear useful in the extension of the current analysis to continuous output channels.

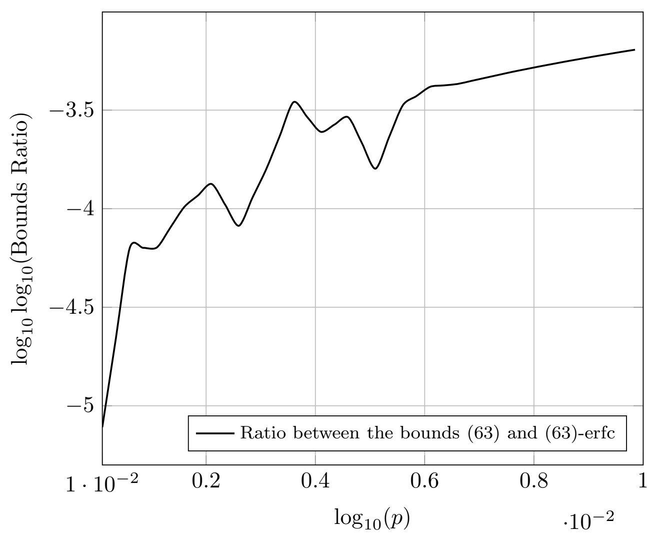
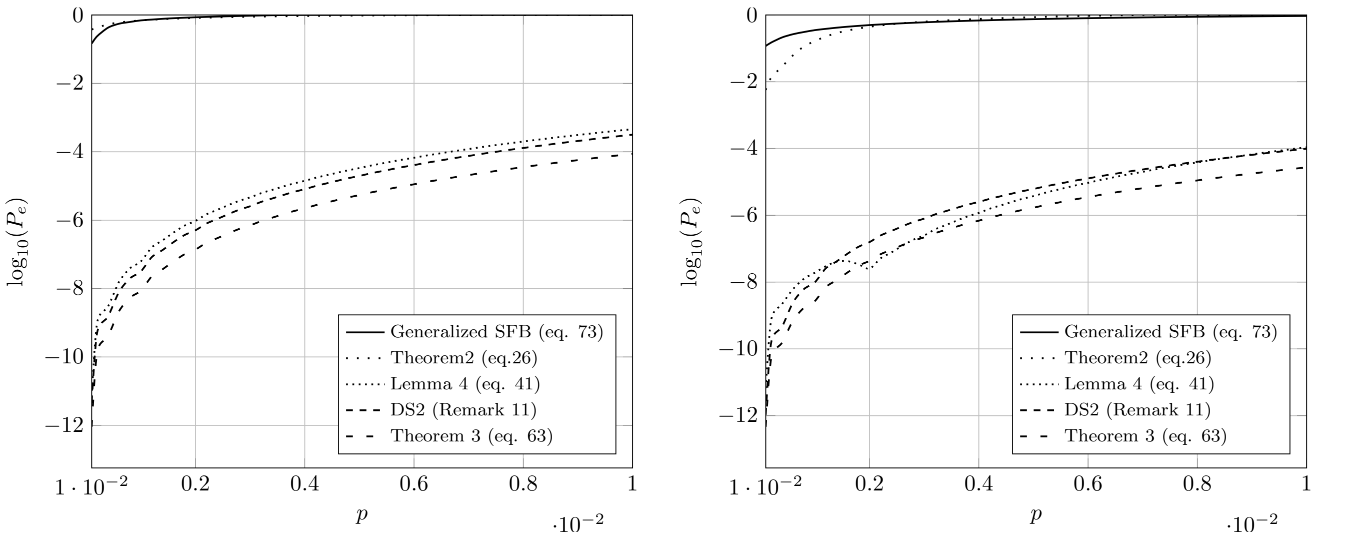


Fig. 5. Ratio between the bounds (63) and (63)-erfc in log log scale for the transmission of the $\mathrm{RM}(2,5)$ code over a binary symmetric channel with error transition probability p. List decoding is performed at the output of the symmetric channel with fixed list size $L = 1$.

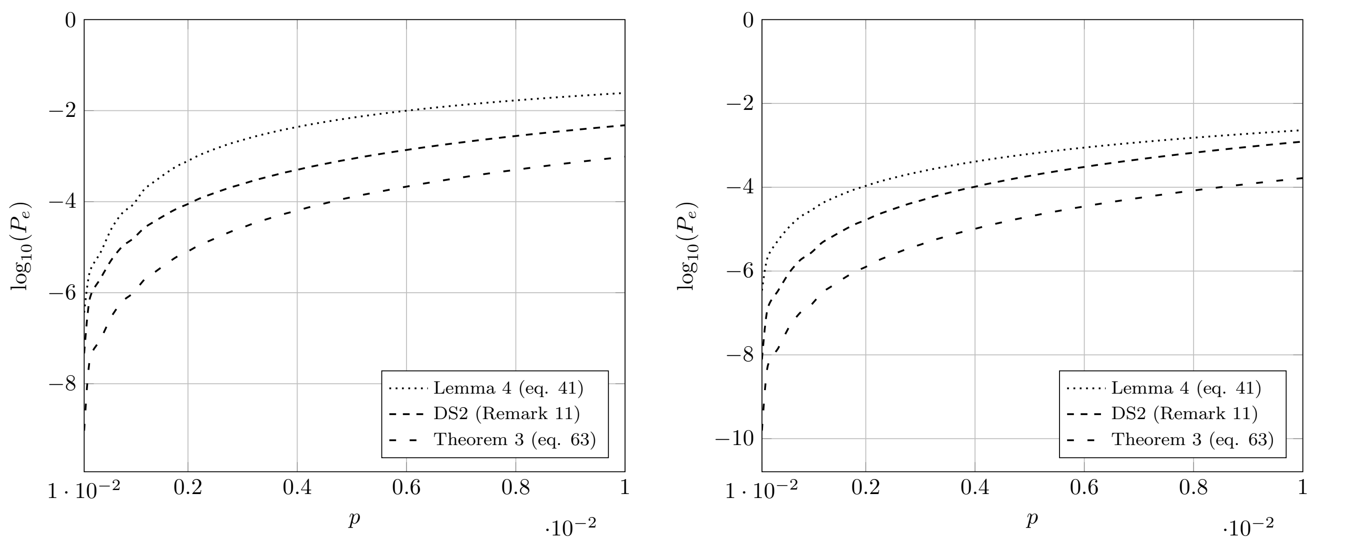
VII. Summary

In the current work, the class of L-list permutation invariant q-ary linear codes is introduced along with a new technique to upper bound their error decoding probability under list decoding. The existence of codes belonging to the defined class is confirmed with the aid of probabilistic arguments, when specific conditions regarding their rate, weight distribution and covering radius are met. Considering a q-ary discrete memoryless output symmetric channel as the transmission



(a) $L_1 = 5$ (b) $L_1 = 16$

Fig. 3. Comparison of the upper bounds (26) of Theorem 2, (41) of Lemma 4 and (63) of Theorem 3 along with the DS2 bound of Remark 11 and generalized Shulman-Feder bound (73) for the transmission of the $\mathrm{RM}(2, 5)$ code over a binary symmetric channel with error transition probability p. List decoding is performed at the output of the symmetric channel with fixed list size L_1.



(a) $L_2 = 1$ (b) $L_2 = 6$

Fig. 4. Comparison of the upper bounds (26) of Theorem 2, (41) of Lemma 4 and (63) of Theorem 3 along with the DS2 bound of Remark 11 and generalized Shulman-Feder bound (73) for the transmission of the $[13, 7, 5]$ Quadratic Residue (QR) code over a 3-ary symmetric channel with error transition probability $p/2$. List decoding is performed at the output of the symmetric channel with fixed list size L_2. For the case $L_2 = 1$, bound (26) of Theorem 2 and the generalized Shulman-Feder bound of (73) are very close to 1 and thus not illustrated in Fig. 4a. For $L_2 = 6$, the generalized Shulman-Feder bound of (73) is close to 1 and thus not depicted in Fig. 4b.

medium, the new bound for the specific class of codes is proved to be tighter than the generalized version of Shulman-Feder bound and relies on the same parameters. A refined version of the new bound is also presented employing Gallager's first bounding technique and separating the error decoding region according to the Hamming weight of the received sequences. Furthermore, the structure of the bound along with the previous existence results provide the reliability function of the treated channels for certain coding rates. The extension of these coding rates to the whole rate region below the cutoff rate of the discrete symmetric channels consists an interesting research goal.

Under proper modifications, a new class of bounding techniques is also derived. The latter class applies to all q-ary linear codes while all its members, characterized by horizontally flip sigmoid functions, improve the DS2 bound as well as all its variations for discrete channels. Moreover, the calculation of the new bounds require exactly the same code information necessary for the calculation of the DS2 bound. The random coding versions of the members of the new class appear to extend the random coding theorem presented by Gallager. Nevertheless, the derivation of analytic expressions for the respective random coding bounds appears rather difficult and requires the analysis of certain properties underpinning the employed horizontally flip sigmoid functions.

The extension of the proposed approach to non-symmetric channels is not straightforward. In particular, the error decoding probability is no longer the same for every message and thus the bounds cannot be calculated for linear codes of high dimension. In addition, the existence of a permutation invariance property for non-symmetric channels appears rather infeasible.

Acknowledgements

The author would like to thank the associate editor for handling the paper and the anonymous reviewers for their helpful comments that greatly enhanced the current work and its presentation.

Appendix A
Proof of Lemma 3

Consider a codeword $\mathbf{c}_m$ of $\mathcal{C}$ of Hamming weight l and let $l_0, l_1, \ldots, l_{q-1}$ denote the number of times the symbols $0, 1, \ldots, q-1$ appear in $\mathbf{c}_m$ respectively. It holds $l_1 + \ldots + l_{q-1} = l$ and $l_0 + l_1 + \ldots + l_{q-1} = N$. Then, the possible number of permutations of $\mathbf{c}_m$ is given by

$$\binom{N}{l_1} \cdot \binom{N-l_1}{l_2} \cdot \ldots \cdot \binom{N-l_1-l_2-\ldots-l_{q-2}}{l_{q-1}} = \frac{N!}{l_0! l_1! \ldots l_{q-1}!} = \binom{N}{l_1, \ldots, l_{q-1}} \tag{76}$$

which is exactly the multinomial coefficient. Consequently, the mean value in the left hand side of (24) is expressed as

$$E\left[\Pr(\mathbf{y}|\mathbf{c}_m, \mathcal{C})^\lambda\right] = \frac{1}{\binom{N}{l_1,\ldots,l_{q-1}}} \sum_{j_1=1}^{\binom{N}{l_1}} \sum_{j_2=1}^{\binom{N-l_1}{l_2}} \cdots \sum_{j_{q-1}=1}^{\binom{N-l_1-\ldots-l_{q-2}}{l_{q-1}}} \Pr\left(\mathbf{y}|\epsilon^{j_1,\ldots,j_{q-1}}\right)^\lambda = E\left[\Pr\left(\mathbf{y}|\epsilon^{l_1,\ldots,l_{q-1}}\right)^\lambda\right] \tag{77}$$

where $\epsilon^{l_1,\ldots,l_{q-1}}$ is a vector of Hamming weight l consisting of $l_1, \ldots, l_{q-1}$ symbols $1, \ldots, q-1$ respectively, and $\epsilon^{j_1,\ldots,j_{q-1}}$ a permutation of $\epsilon^{l_1,\ldots,l_{q-1}}$. It holds

$$\left(\sum_{m\neq 0} E\left[\Pr(\mathbf{y}|\mathbf{c}_m, \mathcal{C})^\lambda\right]\right)^\rho = \left(\sum_{\substack{l=0 \\ l_1+\ldots+l_{q-1}=l}}^{N} S_{l_1,\ldots,l_{q-1}} E\left[\Pr\left(\mathbf{y}|\epsilon^{l_1,\ldots,l_{q-1}}\right)^\lambda\right]\right)^\rho \tag{78}$$

where $S_{l_1,\ldots,l_{q-1}}$ is the number of codewords of $\mathcal{C}$ consisting of $l_1, \ldots, l_{q-1}$ $1, \ldots, q-1$ symbols respectively. Due to the definitions in (25), relation (78) is equivalently expressed as

$$(M-1)^\rho \left(\sum_{\substack{l=0 \\ l_1+\ldots+l_{q-1}=l}}^{N} b_{l_1,\ldots,l_{q-1}} \left(\frac{v_{l_1,\ldots,l_{q-1}}}{b_{l_1,\ldots,l_{q-1}}}\right) \cdot E\left[\Pr\left(\mathbf{y}|\epsilon^{l_1,\ldots,l_{q-1}}\right)^\lambda\right]\right)^\rho . \tag{79}$$

Since

$$\sum_{l=0}^{N} \sum_{l_1+\ldots+l_{q-1}=l} b_{l_1,\ldots,l_{q-1}} = 1 \tag{80}$$

application of the following variant of Hölder's inequality

$$\sum_i P_i \alpha_i \beta_i \leq \left(\sum_i P_i \alpha_i^Q\right)^{\frac{1}{Q}} \left(\sum_i P_i \beta_i^P\right)^{\frac{1}{P}}$$
$$P, Q \geq 0, \quad \frac{1}{Q} + \frac{1}{P} = 1, \quad \sum_i P_i = 1, \quad \alpha_i, \beta_i \geq 0 \tag{81}$$

to (79) leads to the following upper bound

$$(M-1)^\rho \left(\sum_{\substack{l=0 \\ l_1+\ldots+l_{q-1}=l}}^{N} b_{l_1,\ldots,l_{q-1}} \left(\frac{v_{l_1,\ldots,l_{q-1}}}{b_{l_1\ldots,l_{q-1}}}\right)^Q\right)^{\frac{\rho}{Q}} \cdot \left(\sum_{\substack{l=0 \\ l_1+\ldots+l_{q-1}=l}}^{N} b_{l_1,\ldots,l_{q-1}} E\left[\Pr\left(\mathbf{y}|\epsilon^{l_1,\ldots,l_{q-1}}\right)^\lambda\right]^P\right)^{\frac{\rho}{P}} . \tag{82}$$

For the third term of the product in (82) it holds through (77), the definition of $b_{l_1,\ldots,l_{q-1}}$ in (25) and application of Jensen's inequality to the mean value for $P \geq 1$

$$\left(\sum_{\substack{l=0 \\ l_1+\ldots+l_{q-1}=l}}^{N} b_{l_1,\ldots,l_{q-1}} E\left[\Pr\left(\mathbf{y}|\epsilon^{l_1,\ldots,l_{q-1}}\right)^{\lambda P}\right]\right)^{\frac{\rho}{P}} = \left(\sum_{\substack{l=0 \\ l_1+\ldots+l_{q-1}=l}}^{N} \sum_{j_1=0}^{\binom{N}{l_1}} \cdots \sum_{j_{q-1}=0}^{\binom{N-\ldots-l_{q-2}}{l_{q-1}}} \frac{\Pr\left(\mathbf{y}|\epsilon^{j_1,\ldots,j_{q-1}}\right)^{\lambda P}}{q^N}\right)^{\frac{\rho}{P}}$$
$$= \left(q^{-N} \sum_{\mathbf{x}\in F_q^N} \Pr(\mathbf{y}|\mathbf{x})^{\lambda P}\right)^{\frac{\rho}{P}} . \tag{83}$$

If in (82) and (83) we set $\lambda = 1/(1+\rho')P$ and $\rho = \rho' P$ with $\rho' \geq 0$, then the claim of the lemma is established.

Appendix B
Proof of Theorem 2

Since the channel is memoryless we have

$$\left(\sum_{\mathbf{x}\in F_q^N} \Pr(\mathbf{y}|\mathbf{x})^{\frac{1}{1+\rho'}}\right)^{\rho'} = \left(\sum_{x_1\in F_q} \Pr(y_1|x_1)^{\frac{1}{1+\rho'}}\right)^{\rho'} \cdot \ldots \cdot \left(\sum_{x_N\in F_q} \Pr(y_N|x_N)^{\frac{1}{1+\rho'}}\right)^{\rho'} \tag{84}$$

while for every y_i, due to the channel's symmetry, it holds

$$\left(\sum_{x_i\in F_q} \Pr(y_i|x_i)^{\frac{1}{1+\rho'}}\right)^{\rho'} = \left[\hat{p}^{\frac{1}{1+\rho'}} + (q-1)\hat{p}_q^{\frac{1}{1+\rho'}}\right]^{\rho'} \tag{85}$$

where $\hat{p} = 1-p$ and $\hat{p}_q = p/(q-1)$. Thus from (84) and (85)

$$\left(\sum_{\mathbf{x}\in F_q^N} \Pr(\mathbf{y}|\mathbf{x})^{\frac{1}{1+\rho'}}\right)^{\rho'} = \left[\hat{p}^{\frac{1}{1+\rho'}} + (q-1)\hat{p}_q^{\frac{1}{1+\rho'}}\right]^{N\rho'}. \quad (86)$$

Moreover, due to the memoryless nature of the channel it holds

$$\Pr(\mathbf{y}|\mathbf{c}_0) = \prod_{i=1}^{N} \Pr(y_i|0) = \hat{p}^{N-wt(\mathbf{y})}\hat{p}_q^{wt(\mathbf{y})}. \quad (87)$$

If we substitute (86) to (24) of Lemma 3 and apply the result along with (87) to the sum in (23), we deduce that each $\mathbf{y}$ of Hamming weight l contributes to the previous sum with the term

$$f_{\text{de}}\left(\frac{q^{N\rho'}L^{\rho'P}}{(M-1)^{\rho'P}} \frac{\left[\hat{p}^{N-l}\hat{p}_q^l\right]^{\frac{\rho'}{1+\rho'}}}{\mathcal{F}_{\mathcal{C}}(P)^{\rho'}\left[\hat{p}^{\frac{1}{1+\rho'}} + (q-1)\hat{p}_q^{\frac{1}{1+\rho'}}\right]^{\rho'N}}\right) \quad (88)$$

where, under the relation $1/P + 1/Q = 1$ of (25), $\mathcal{F}_{\mathcal{C}}(P)$ is given by (28). Since the number of vectors $\mathbf{y}$ with Hamming weight l consisting of l_i symbols $i \in [1, q-1]$ is provided by the multinomial coefficient

$$\binom{N}{l_1, \ldots, l_{q-1}} \quad (89)$$

where it is noted

$$\sum_{l_1+\ldots+l_{q-1}=l} \binom{N}{l_1, \ldots, l_{q-1}} = \binom{N}{l}(q-1)^l \quad (90)$$

the upper bound in (23) is expressed with the aid of (88)-(90) as

$$P_e^{\mathcal{L}} \le \sum_{l=0}^{N}\binom{N}{l}\hat{p}^{N-l}p^l f_{\text{de}}\left(\frac{q^{N\rho'}L^{\rho'P}}{(M-1)^{\rho'P}}\cdot \frac{\left[\hat{p}^{N-l}\hat{p}_q^l\right]^{\frac{\rho'}{1+\rho'}}}{\mathcal{F}_{\mathcal{C}}(P)^{\rho'}\left[\hat{p}^{\frac{1}{1+\rho'}} + (q-1)\hat{p}_q^{\frac{1}{1+\rho'}}\right]^{\rho'N}}\right). \quad (91)$$

We next proceed to optimize the upper bound of (91) under the assumptions $L < S_{l_0^*,\ldots,l_{q-1}^*}$ and (29) of Remark 8 by minimizing the corresponding function

$$g(P) \triangleq P\ln\frac{M-1}{L} + \ln\mathcal{F}_{\mathcal{C}}(P) \quad (92)$$

with respect to $P \ge 1$. To simplify the analysis, we consider without loss of generality the binary alphabet case $F_2 = \{0,1\}$ so that $g(P)$ gets the form

$$g(P) = P\ln\frac{M-1}{L} + (P-1)\ln\sum_{l=0}^{N} b_l\left(\frac{v_l}{b_l}\right)^{\frac{P}{P-1}}. \quad (93)$$

The sign of the second derivative of $g(P)$ in (93) is determined by the sign of the following term

$$\left(\sum_{l=0}^{N} b_l\left(\frac{v_l}{b_l}\right)^{\frac{P}{P-1}}\left(\ln\frac{v_l}{b_l}\right)^2\right)\left(\sum_{l=0}^{N} b_l\left(\frac{v_l}{b_l}\right)^{\frac{P}{P-1}}\right) - \left(\sum_{l=0}^{N} b_l\left(\frac{v_l}{b_l}\right)^{\frac{P}{P-1}}\ln\frac{v_l}{b_l}\right)^2$$

which is equivalent to an expression of the form

$$\left(\sum_i c_i d_i^2\right)\left(\sum_i c_i\right) - \left(\sum_i c_i d_i\right)^2, \quad c_i \ge 0.$$

The latter expression is easily proved to be a non-negative quantity so that $g(P)$ of (93) is a convex function of $P \ge 1$. Indeed, our argument is justified if for positive sequences α_i and β_i we consider Hölder's inequality

$$\left(\sum_i \alpha_i\beta_i\right)^2 \le \sum_i \alpha_i^2 \sum_i \beta_i^2$$

and set $\beta_i = \sqrt{c_i}$ and $\alpha_i = \sqrt{c_i}|d_i|$. An equivalent representation of (93) is

$$g(P) = P\ln\frac{M-1}{L} + (P-1)\ln\sum_{l=0}^{N} v_l\left(\frac{v_l}{b_l}\right)^{\frac{1}{P-1}}. \quad (94)$$

The first derivative of $g(P)$ is an increasing function of P since $g(P)$ is convex and equals

$$\frac{dg(P)}{dP} = \ln\frac{M-1}{L} + \ln\sum_{l=0}^{N} v_l\left(\frac{v_l}{b_l}\right)^{\frac{1}{P-1}} - \frac{1}{P-1}\frac{\sum_{l=0}^{N} v_l\left(\frac{v_l}{b_l}\right)^{\frac{1}{P-1}}\ln\frac{v_l}{b_l}}{\sum_{l=0}^{N} v_l\left(\frac{v_l}{b_l}\right)^{\frac{1}{P-1}}}.$$

Furthermore, as P approaches 1 from above it holds

$$\lim_{P\to 1^+}\frac{dg(P)}{dP} = \ln\frac{M-1}{L} + \ln v_{l^*} = \ln\frac{S_{l^*}}{L}$$
$$\text{where} \quad l^* = \arg\max_l \frac{v_l}{b_l}.$$

According to the assumption of Remark 8, reduced to $L < S_{l^*}$ in the binary case, the above limit is positive and thus due to convexity the first order derivative of $g(P)$ is positive. Consequently, $g(P)$ is convex and monotonic increasing and its minimum value is achieved as P approaches 1. The limit of $g(P)$ as P approaches 1 from above is equal to $\mathcal{F}_{\mathcal{C}}$ defined in (30).

Appendix C
Extension of Theorem 3 to the general alphabet case

Since the channel is memoryless, each factor in the double exponent of (62) satisfies

$$\frac{\Pr(\mathbf{y}|\mathbf{c}_0)}{\Pr(\mathbf{y}|\mathbf{c}_m)} = \prod_{i_1\in I_1^m}\frac{\Pr(y_{i_1}|0)}{\Pr(y_{i_1}|1)}\prod_{i_2\in I_2^m}\frac{\Pr(y_{i_2}|0)}{\Pr(y_{i_2}|2)}\cdot\ldots\cdot \prod_{i_{q-1}\in I_{q-1}^m}\frac{\Pr(y_{i_{q-1}}|0)}{\Pr(y_{i_{q-1}}|q-1)} = \prod_{j=1}^{q-1}\prod_{i_j\in I_j^m}\frac{\Pr(y_{i_j}|0)}{\Pr(y_{i_j}|j)} \quad (95)$$

where each subset I_j^m, $j \in [0, q-1]$ with cardinality $|I_j^m|$ consists of the positions where symbol j appears in the codeword $\mathbf{c}_m$ respectively. For every fixed $\mathbf{y} \in F_q^N$ we analyze the products in (95). Let the set Y_{k,I_j^m} consist of the positions of vector $\mathbf{y}$ where symbol $k \in F_q$ appears inside the respective subset I_j^m of $\mathbf{y}$. Since the considered channel is also output symmetric, each product of (95) satisfies according to the previous notation for $j \geq 1$

$$\prod_{i_j \in I_j^m} \frac{\Pr(y_{i_j}|0)}{\Pr(y_{i_j}|j)} = \left(\frac{\hat{p}_q}{\hat{p}}\right)^{|Y_{j,I_j^m}|} \left(\frac{\hat{p}}{\hat{p}_q}\right)^{|Y_{0,I_j^m}|} = \left(\frac{\hat{p}_q}{\hat{p}}\right)^{|Y_{j,I_j^m}|-|Y_{0,I_j^m}|}. \quad (96)$$

Embedding (96) into (95) we have

$$\frac{\Pr(\mathbf{y}|\mathbf{c}_0)}{\Pr(\mathbf{y}|\mathbf{c}_m)} = \left(\frac{\hat{p}_q}{\hat{p}}\right)^{\sum_{j=1}^{q-1}\left(|Y_{j,I_j^m}|-|Y_{0,I_j^m}|\right)}. \quad (97)$$

The previous product of terms is independent of Y_{0,I_0^m} as well as Y_{k,I_j^m} for $k \neq j$. Moreover, it is noted that $|I_0^m| = N - \text{wt}(\mathbf{c}_m)$ and $\sum_{j=1}^{q-1} |I_j^m| = \text{wt}(\mathbf{c}_m)$. For the binary alphabet case where $q = 2$ we have

$$\sum_{j=1}^{q-1}\left(|Y_{j,I_j^m}| - |Y_{0,I_j^m}|\right) = |Y_{1,I_1^m}| - |Y_{0,I_1^m}|$$

and

$$\text{wt}(\mathbf{c}_m + \mathbf{c}_0) = |Y_{1,I_1^m}| + |Y_{0,I_1^m}|$$

so that

$$\sum_{j=1}^{q-1}\left(|Y_{j,I_j^m}| - |Y_{0,I_j^m}|\right) = \text{wt}(\mathbf{c}_m) - 2|Y_{0,I_1^m}|.$$

Furthermore, it holds $|Y_{0,I_1^m}| = \text{wt}(\mathbf{c}_m) - \text{wt}(\mathbf{y} \cdot \mathbf{c}_m)$ so that combining all previous relations we deduce

$$\sum_{j=1}^{q-1}\left(|Y_{j,I_j^m}| - |Y_{0,I_j^m}|\right) = 2\text{wt}(\mathbf{y} \cdot \mathbf{c}_m) - \text{wt}(\mathbf{c}_m) = \text{wt}(\mathbf{y}) - \text{wt}(\mathbf{y} + \mathbf{c}_m).$$

Consider a subsequence $\{y_{i_j}\}_{i_j \in I_j^m}$ and suppose that symbol j appears in positions Y_{j,I_j^m} over it while symbol 0 appears in positions Y_{0,I_j^m}. The rest of the positions of the subsequence $\{y_{i_j}\}_{i_j \in I_j^m}$, in total $|I_j^m| - |Y_{j,I_j^m}| - |Y_{0,I_j^m}|$, are occupied by symbols from the subset $A_j \triangleq \{0, 1, \ldots, q-1\} \setminus \{0, j\}$, with cardinality $|A_j| = q - 2$. Since with $q - 2$ symbols from A_j we are able to construct

$$\sum_{k_1=0}^{J} \cdots \sum_{k_{j-1}=0}^{J-\sum_{i=1}^{j-2} k_i} \sum_{k_{j+1}=0}^{J-\sum_{i=1}^{j-1} k_i} \cdots \sum_{k_{q-2}=0}^{J-\sum_{i\in A_j} k_i} \binom{J}{k_1, \ldots, k_{j-1}, k_{j+1}, \ldots, k_{q-2}} = (q-1)^J \quad (98)$$

sequences, the total number of subsequences $\{y_{i_j}\}_{i_j \in I_j^m}$ with $|Y_{j,I_j^m}|$ symbols j and $|Y_{0,I_j^m}|$ symbols 0 is given by

$$\binom{|I_j^m|}{|Y_{j,I_j^m}|, |Y_{0,I_j^m}|}(q-1)^{|I_j^m|-|Y_{j,I_j^m}|-|Y_{0,I_j^m}|}. \quad (99)$$

It is reminded that

$$\binom{|I_j^m|}{|Y_{j,I_j^m}|, |Y_{0,I_j^m}|} = \binom{|I_j^m|}{|Y_{j,I_j^m}|}\binom{|I_j^m| - |Y_{j,I_j^m}|}{|Y_{0,I_j^m}|}$$

is the multinomial coefficient defined in (76). The main reason for expressing the possible number of subsequences of length J, whose elements take values in A_j, with the multiple sums of the left hand side of (98) is our effort to designate the number k_i of times symbols $i \in A_j$ appear in the respective subsequences.

Let the un-normalized tilting measure $G_0(\mathbf{y}) = \prod_{i=1}^N g_0(y_i)$ for some positive function $g_0(y)$ and $\mathcal{N}$ the set of natural numbers $[1, N]$. Under the previous assumption and the fact that (97) does not depend on Y_{0,I_0^m}, the inner summation of the second product in the right hand side of (62) is equivalently expressed as

$$\sum_{\mathbf{y} \in F_q^N} \Pr(\mathbf{y}|\mathbf{c}_0) G_0(\mathbf{y})^{1-\frac{1}{\rho}} f_{\text{de}}\left(\left(\frac{\Pr(\mathbf{y}|\mathbf{c}_0)}{\Pr(\mathbf{y}|\mathbf{c}_m)}\right)^{\lambda}\right) = \prod_{i \in I_0^m}\left(\sum_{y_i \in F_q} \Pr(y_i|0) g_0(y_i)^{1-\frac{1}{\rho}}\right) \cdot \left(\sum_{k \in \mathcal{N}\setminus I_0^m} \sum_{y_k \in F_q}\left(\prod_{j=1}^{q-1}\prod_{i_j \in I_j^m} \Pr(y_{i_j}|0) g_0(y_{i_j})^{1-\frac{1}{\rho}}\right) \cdot f_{\text{de}}\left(\prod_{j=1}^{q-1}\prod_{i_j \in I_j^m} \frac{\Pr(y_{i_j}|0)^{\lambda}}{\Pr(y_{i_j}|j)^{\lambda}}\right)\right). \quad (100)$$

For the first term of the product in the right hand side of (100), since $|I_0^m| = N - \text{wt}(\mathbf{c}_m)$ it holds

$$\prod_{i \in I_0^m}\left(\sum_{y_i \in F_q} \Pr(y_i|0) g_0(y_i)^{1-\frac{1}{\rho}}\right) = \left(\sum_{y \in F_q} \Pr(y|0) g_0(y)^{1-\frac{1}{\rho}}\right)^{N-\text{wt}(\mathbf{c}_m)}. \quad (101)$$

For the second term of the same product, the summation is noted to be over only those $y_i \in F_q$ where $i \in \mathcal{N} \setminus I_0^m$. We express the subsequence $\{y_i\}_{i \in \mathcal{N}\setminus I_0^m}$ as the concatenation of subsets $\{y_i\}_{i \in I_1^m}, \{y_i\}_{i \in I_2^m}, \ldots, \{y_i\}_{i \in I_{q-1}^m}$ and consider a possible value of that subsequence. For this value the respective summand of the product in the right hand side of (100) equals in accordance to (97) and the discussion justifying relations (98) and (99)

$$\hat{p}^{\sum_{j=1}^{q-1}|Y_{0,I_j^m}|} \hat{p}_q^{\sum_{j=1}^{q-1}|Y_{j,I_j^m}|} g_0(0)^{\sum_{j=1}^{q-1}|Y_{0,I_j^m}|} \prod_{j=1}^{q-1} g_0(j)^{|Y_{j,I_j^m}|} \cdot f_{\text{de}}\left(\left(\frac{\hat{p}_q}{\hat{p}}\right)^{\lambda \sum_{j=1}^{q-1}\left(|Y_{j,I_j^m}|-|Y_{0,I_j^m}|\right)}\right) \quad (102)$$

where it is reminded $\hat{p} = 1 - p$ and $\hat{p}_q = p/(q-1)$. Using the notation of subset A_j for $j \in [1, q-1]$ defined in the paragraph before (98), the previous equation is expressed equivalently as the product of the following two terms

$$f_1\left(\lambda,\rho,\{|Y_{0,I_j^m}|\}_{j=1}^{q-1},\{|Y_{j,I_j^m}|\}_{j=1}^{q-1}\right)\triangleq \hat{p}^{\sum_{j=1}^{q-1}|Y_{0,I_j^m}|}$$
$$\hat{p}_q^{\sum_{j=1}^{q-1}|Y_{j,I_j^m}|}g_0(0)^{(1-\frac{1}{\rho})\sum_{j=1}^{q-1}|Y_{0,I_j^m}|}\prod_{j=1}^{q-1}g_0(j)^{(1-\frac{1}{\rho})|Y_{j,I_j^m}|}$$
$$f_{\text{de}}\left(\left(\frac{\hat{p}_q}{\hat{p}}\right)^{\lambda\sum_{j=1}^{q-1}\left(|Y_{j,I_j^m}|-|Y_{0,I_j^m}|\right)}\right) \tag{103}$$

and

$$\prod_{j=1}^{q-1}\left(\hat{p}_q g_0(j)^{1-\frac{1}{\rho}}\right)^{\sum_{i\in A_j}|Y_{j,I_i^m}|}. \tag{104}$$

The two terms above are noted to be independent regarding the subscripts of the subsequence $\{y_i\}_{i\in\mathcal{N}\setminus I_0^m}$. Indeed, the first term (103) depends only on the subscripts $\bigcup_{j=1}^{q-1}\{Y_{0,I_j^m}\cup Y_{j,I_j^m}\}$ while the second term (104) depends on the rest positions $\{\mathcal{N}\setminus I_0^m\}\setminus\bigcup_{j=1}^{q-1}\{Y_{0,I_j^m}\cup Y_{j,I_j^m}\}$. Thus, if we consider the summation on the second term of the product in the right hand side of (100), for a subsequence $\{y_i\}_{i\in\mathcal{N}\setminus I_0^m}$ whose elements in positions $\bigcup_{j=1}^{q-1}\{Y_{0,I_j^m}\cup Y_{j,I_j^m}\}$ remain fixed, and thus the term (103) is fixed, we get the following sum of terms of the form in (104) over all possible combinations over $\{\mathcal{N}\setminus I_0^m\}\setminus\bigcup_{j=1}^{q-1}\{Y_{0,I_j^m}\cup Y_{j,I_j^m}\}$ with the help of the expansion in the left hand side of (98)

$$f_2\left(\rho,\{|I_j^m|\}_{j=1}^{q-1},\{|Y_{j,I_j^m}|\}_{j=1}^{q-1},\{|Y_{0,I_j^m}|\}_{j=1}^{q-1}\right)\triangleq$$
$$\sum_{j=1}^{q-1}\sum_{\sum_{k\neq j}l_k^j=X_j}\binom{X_j}{l_1^j,\ldots,l_{j-1}^j,l_{j+1}^j,\ldots,l_{q-1}^j}.$$
$$\prod_{k=1}^{q-1}\left(\hat{p}_q g_0(k)^{1-\frac{1}{\rho}}\right)^{\sum_{i\in A_k}l_k^i} \tag{105}$$

where $X_j=|I_j^m|-|Y_{j,I_j^m}|-|Y_{0,I_j^m}|$. The latter term is analogous to the one given in (69) in the proof of Theorem 3. If we group together all subsequences $\{y_i\}_{i\in\mathcal{N}\setminus I_0^m}$ with the same values regarding the sets $\{|Y_{0,I_j^m}|\}_{j=1}^{q}$ and $\{|Y_{j,I_j^m}|\}_{j=1}^{q}$, then the second term in the right hand side of (100) equals with the aid of (103) and (105)

$$f_3\left(\lambda,\rho,\{|I_j^m|\}_{j=1}^{q-1}\right)\triangleq\sum_{j=1}^{q-1}\sum_{i=0}^{q-1}\sum_{|Y_{i,I_j^m}|=0}^{|I_j^m|-\sum_{k=0}^{i-1}|Y_{k,I_j^m}|}$$
$$\prod_{j=1}^{q-1}\prod_{i=0}^{q-1}\binom{|I_j^m|-\sum_{k=0}^{i-1}|Y_{k,I_j^m}|}{|Y_{i,I_j^m}|}.$$
$$f_1\left(\lambda,\rho,\{|Y_{0,I_j^m}|\}_{j=1}^{q-1},\{|Y_{j,I_j^m}|\}_{j=1}^{q-1}\right)\cdot$$
$$f_2\left(\rho,\{|I_j^m|\}_{j=1}^{q},\{|Y_{j,I_j^m}|\}_{j=1}^{q-1},\{|Y_{0,I_j^m}|\}_{j=1}^{q-1}\right). \tag{106}$$

Substituting (106) and (101) into the right hand side of (100), the bound in (63) is expressed equivalently as shown in (74), where it is reminded that $l_1,\ldots,l_{q-1}$ is the number of codewords in $\mathcal{C}$ that consist of l_i symbols i respectively.

REFERENCES

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

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

[3] N. Shulman and M. Feder, "Random coding techniques for nonrandom codes," *IEEE Trans. Inf. Theory*, vol. 45, pp. 2101–2014, Sept. 1999.

[4] D. Divsalar, "A simple tight bound on error probability of block codes with application to turbo codes," JPL, TMO Progr. Rep. 42-139, Nov. 1999.

[5] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inf. Theory*, vol. IT-11, pp. 3–18, 1965.

[6] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels I-II," *Information and Control*, vol. 10, pp. 65–103, 522–552, 1967.

[7] A. Valembois and M. Fossorier, "Sphere-packing bounds revisited for moderate block length," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 2998–3014, Dec. 2004.

[8] G. Wiechman and I. Sason, "An improved sphere-packing bound for finite-length codes over symmetric memoryless channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1962–1990, May 2008.

[9] G. D. Forney, "Exponential error bounds for erasure,list, and decision feedback schemes," *IEEE Trans. Inf. Theory*, vol. 14, no. 2, pp. 206–220, Mar. 1968.

[10] N. Merhav, "List decoding-random coding exponents and expurgated exponents," *IEEE Trans. Inf. Theory*, Nov. 2013, submitted.

[11] I. E. Bocharova, R. Johannesson, B. D. Kudryashov, and M. Loncar, "An improved bound on the list error probability and list distance properties," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 13–32, Jan. 2008.

[12] E. Hof, I. Sason, and S. Shamai, "Performance bounds for erasure, list and feedback schemes with linear block codes," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3754–3778, Aug. 2010.

[13] N. Alon and J. H. Spencer, *The Probabilistic Method, 3rd edition*. John Wiley & Sons, Inc., Sep. 2007.

[14] K. Xenoulis and N. Kalouptsidis, "Tight performance bounds for permutation invariant binary linear block codes over symmetric channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 9, pp. 6015–6024, Sep. 2011.

[15] R. G. Gallager and B. Nakiboglu, "Variations on a theme by Schalkwijk and Kailath," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 6–17, Jan. 2010.

[16] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. Wiley-Interscience, 2006.

[17] T. Baicheva, S. Dodunekov, and R. Koetter, "On the performance of the ternary [13,7,5] quadratic-residue code," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 562–564, Feb. 2002.

[18] C. Huffman and V. Pless, *Fundamentals of Error Correcting Codes*. Cambridge University Press, 2003.

[19] V. Guruswami, J. Hastad, and S. Kopparty, "On the list-decodability of random linear codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 718–725, Feb. 2011.

[20] V. Guruswami, *List Decoding of Error-Correcting Codes: Winning Thesis of the 2002 ACM Doctoral Dissertation*. Springer-Verlag New York, Inc., 2005.

[21] G. Como and F. Fagnani, "The capacity of finite abelian group codes over symmetric memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2037–2054, May 2009.

[22] A. Barg and G. D. Forney, Jr., "Random codes: minimum distances and error exponents," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2568–2573, Sept. 2002.

[23] A. Ashikhmin, A. Barg, and S. Litsyn, "Estimates of the distance distribution of nonbinary codes, with applications," in *Codes and Association Schemes*, A. Barg and S. Litsyn, Eds. Providence, RI: AMS, 2001.

[24] G. Cohen, M. Karpovsky, H. Mattson, and J. Schatz, "Covering radius-survey and recent results," *IEEE Trans. Inf. Theory*, vol. 31, no. 3, pp. 328–343, May 1985.

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

[26] M. Twitto, I. Sason, and S. Shamai, "Tightened upper bounds on the ML decoding error probability of binary linear block codes," *IEEE Trans. Inf. Theory*, vol. 53, no. 4, pp. 1495–1510, Apr. 2007.

[27] K. Xenoulis and N. Kalouptsidis, "Refinement of the DS2 bound and its extensions for discrete memoryless symmetric channels," *IEEE Commun. Letters*, vol. 18, no. 5, pp. 861–864, May 2014.

[28] R. G. Gallager, "The random coding bound is tight for the average code," *IEEE Trans. Inf. Theory*, pp. 244–246, 1973.

Kostis Xenoulis received the Ph.D. degree from the Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Greece, in 2010. He is currently a postdoctoral researcher at the same department. His research interests are in the wide area of information theory and coding.