Variations on the Gallager Bounds, Connections and Applications*

Shlomo Shamai
Department of Electrical Engineering
Technion – Israel Institute of Technology
Haifa 32000, Israel

Igal Sason
Department of Communications and Computer Science
EPFL – Swiss Federal Institute of Technology
Lausanne, CH–1015, Switzerland

December 3, 2002

Abstract

In recent years there has been renewed interest in deriving tight bounds on the error performance of specific codes and ensembles, based on their distance spectrum. In this paper, we discuss many reported upper bounds on the maximum likelihood decoding error probability and demonstrate the underlying connections that exist between them. In addressing the Gallager bounds and their variations, we focus on the Duman and Salehi variation, which originates from the standard Gallager bound. A large class of efficient recent bounds (or their Chernoff versions) is demonstrated to be a special case of the generalized second version of the Duman and Salehi bounds. Implications and applications of these observations are pointed out, including the fully interleaved fading channel, resorting to either matched or mismatched decoding. The proposed approach can be generalized to geometrically uniform non-binary codes, finite state channels, bit interleaved coded modulation systems, and it can be also used for the derivation of upper bounds on the conditional decoding error probability.

Keywords: Decoding error probability, distance spectrum, fading channels, Gallager bounds, maximum-likelihood decoding, mismatched decoding, upper bounds.

*This work was supported by the Neaman Institute of Technion and the Israeli Software Radio Consortium. The material in this paper was presented in part at the Second International Symposium on Turbo Codes & Related Topics, Brest, France, 4–7 September 2000, and was also presented in part at the 2001 IEEE International Symposium on Information Theory (ISIT 2001), Washington, D.C., USA, June 24–29, 2001. The research work was done while Igal Sason was a Ph.D. student in the department of electrical engineering at the Technion – Israel Institute of Technology, Haifa 32000, Israel.

1. Introduction

Since the error performance of coded communication systems rarely admits exact expressions, tight analytical bounds serve as a useful theoretical and engineering tool for assessing performance and for gaining insight into the effect of the main system parameters. As specific good codes are hard to identify, the performance of ensembles of codes is usually considered. Fano [14] and Gallager [15] upper bounds were introduced as efficient tools to determine the error exponents of the ensemble of random codes, providing informative results up to the ultimate capacity limit. Since the advent of information theory, the search for efficient coding systems has motivated the introduction of efficient bounding techniques tailored to specific codes or some carefully chosen ensembles of codes. A classical example is the adaptation of the Fano upper bounding technique [14] to specific codes, as reported in the seminal dissertation by Gallager [17] (to be referred to as the Gallager-Fano 1961 bound). The incentive for introducing and applying such bounds has increased with the recent introduction of turbo codes [5] and the rediscovery of low-density parity-check (LDPC) codes [21]. Clearly, the desired bounds must not be subject to the union bound limitation, since for long blocks these ensembles of turbo-like codes perform reliably at rates which are considerably above the cutoff rate (R_0) of the channel (recalling that union bounds for long codes are not informative at the portion of the rate region above R_0, where the performance of these capacity-approaching codes is most appealing). Although maximum likelihood (ML) decoding is in general prohibitively complex for long codes, the derivation of upper bounds on the ML decoding error probability is of interest, providing an ultimate indication of the system performance. Further, the structure of efficient codes is usually not available, necessitating efficient bounds on performance to rely only on basic features, such as the distance spectrum and input-output weight enumeration function (IOWEF) of the examined codes. These latter features can usually be found by some analytical methods (e.g, [22]).

In classical treatments, due to the difficulty in the analytic characterization of optimal codes, random codes were introduced ([14], [15], [16]). This is still the case with modern approaches and practical coding techniques, where ensembles of codes (e.g., turbo [5] and LDPC codes [17]) lend themselves to analytical treatment, while this is not necessarily the case for specifically chosen codes within these families. A desirable feature is to identify efficient bounding techniques encompassing both specific codes and ensembles.

In our work, we focus on the second version of the recently introduced bounds by Duman and Salehi ([11], [12]), whose derivation is based on the 1965 Gallager bounding technique ([15], [16]). Though originally derived for binary signaling over an additive white Gaussian noise (AWGN) channel, we demonstrate here its considerable generality and show that it provides the natural bridge between the 1961 and 1965 Gallager bounds ([15], [17]). It is suitable for both random and specific codes

[7], as well as for either bit or block error probability analysis. It is also demonstrated here that a large class of efficient recent bounds or their Chernoff versions are special cases of the generalized second version of the Duman and Salehi (DS2) bound. We exemplify the use of this generalized bound in various settings, such as the fully interleaved fading channel ([26],[27]).

In an important recent contribution, Divsalar [7] has introduced some efficient and easily applicable bounds, and has also provided insightful observations on the Duman and Salehi bounding technique ([11], [12]) in view of other bounds. In our setting, we shall use and build on some of the interesting observations in [7].

The paper is organized as follows: The 1965 Gallager bound ([15], [16]) and the DS2 bound [12] are presented as preliminary material in Section 2. These two upper bounds form the underlying bounding technique in this paper, as we rely on them throughout. In Section 3, the 1961 Gallager-Fano bound [17] is presented, and some interconnections among the Gallager bounds and the DS2 bound are demonstrated. It is shown in Section 3 that the DS2 bound provides the natural bridge between the 1961 and 1965 Gallager bounds. In Section 4, it is demonstrated that many reported bounds on the ML decoding error probability (which were originally derived independently) can be considered as special cases of the DS2 bound. The 1965 Gallager random coding bound is extended in Section 5 to the mismatched decoding regime, where the decoder operates in a ML fashion, but may use a mismatched metric. These Gallager-type bounds which are derived for the mismatched decoding regime can be applied to deterministic codes and ensembles. Some reported results ([18], [19], [23]) are derived in Section 5, based on an alternative approach which appropriately limits the code ensemble. Some applications and examples of these bounds are presented in Section 6, which include fully interleaved fading channels and mismatched metrics. Finally, Section 7 concludes the paper and three appendices provide supplementary technical details.

2. Preliminaries (Basic Bounds)

2.1. The 1965 Gallager bound

It is well known that the ML decoding error probability conditioned on an arbitrary transmitted (length-N) codeword $\underline{x}^m$ ($P_{e|m}$), is upper bounded by the 1965 Gallager bound ([15], [16]):

$$P_{e|m} \le \sum_{\underline{y}} p_N(\underline{y}|\underline{x}^m) \left(\sum_{m' \neq m} \left(\frac{p_N(\underline{y}|\underline{x}^{m'})}{p_N(\underline{y}|\underline{x}^m)} \right)^{\lambda} \right)^{\rho} , \quad \lambda,\, \rho \ge 0 \,. \tag{1}$$

Here, $\underline{y}$ designates the observation vector (of N components) and $p_N(\underline{y}|\underline{x})$ is the channel's transition probability measure. The upper bound (1) is usually not easily evaluated in terms of basic features of particular codes, except for example, orthogonal codes and the special case of $\rho = 1$ and $\lambda = \frac{1}{2}$ (which yields the Bhattacharyya-union bound).

For discrete, memoryless and output-symmetric channels, the upper bound (1) is not directly applicable for actual code performance calculation, since it cannot be factored into single-letter expressions (because of the ρ^{th} power which operates on the *inner* summation of (1)). Therefore, the bound does not lend itself to code performance calculation in terms of the distance spectrum of the ensemble of codes. This difficulty could be circumvented had the ρ^{th} power in (1) been taken over the expression $\sum_{\underline{y}} \left\{ p_N(\underline{y}|\underline{x}^m) \sum_{m' \neq m} \left(\frac{p_N(\underline{y}|\underline{x}^{m'})}{p_N(\underline{y}|\underline{x}^m)} \right)^{\lambda} \right\}$ (instead of its actual location in the inner summation of the bound (1)).

For ensembles of random block codes, where the codewords are independently selected with an arbitrary probability distribution $q_N(\underline{x})$, Gallager derived an upper bound on the average ML decoding error probability (where the average is over the randomly and independently selected codewords). By invoking the Jensen inequality in (1) ($E[x^\rho] \leq (E[x])^\rho$ where $0 \leq \rho \leq 1$), and setting $\lambda = \frac{1}{1+\rho}$, the 1965 Gallager random coding bound is obtained:

$$P_e \leq (M-1)^\rho \sum_{\underline{y}} \left(\sum_{\underline{x}} q_N(\underline{x})\, p_N(\underline{y}|\underline{x})^{\frac{1}{1+\rho}} \right)^{1+\rho} \quad , \quad 0 \leq \rho \leq 1 \,. \tag{2}$$

Here P_e designates the average decoding error probability, and M is the number of the codewords. For the particular case of a memoryless channel $\left(p_N(\underline{y}|\underline{x}) = \prod_{i=1}^{N} p(y_i|x_i) \right)$ and a memoryless input distribution $\left(q_N(\underline{x}) = \prod_{i=1}^{N} q(x_i) \right)$, the Gallager random coding bound (2) admits the form:

$$P_e \leq e^{-N \cdot E(R,q)} \,, \tag{3}$$

where $R = \frac{\ln M}{N}$ is the code rate (in nats per channel use) and the associated error exponent is

$$E(R,q) \;=\; \max_{0 \leq \rho \leq 1} \left\{ -\ln \left(\sum_{y} \left(\sum_{x} q(x)\, p(y|x)^{\frac{1}{1+\rho}} \right)^{1+\rho} \right) - \rho R \right\} . \tag{4}$$

2.2. The DS2 bound

The Duman and Salehi bounding technique ([11],[12]) originates from the 1965 Gallager bound: Let $\psi_N^m(\underline{y})$ be an arbitrary probability measure (which may also depend on the transmitted codeword

$\underline{x}^m$). The 1965 Gallager bound (1) then yields:

$$\begin{aligned} P_{e|m} &\leq \sum_{\underline{y}} \psi_N^m(\underline{y}) \cdot \psi_N^m(\underline{y})^{-1} \cdot p_N(\underline{y}|\underline{x}^m) \cdot \left(\sum_{m' \neq m} \left(\frac{p_N(\underline{y}|\underline{x}^{m'})}{p_N(\underline{y}|\underline{x}^m)} \right)^{\lambda} \right)^{\rho} \\ &= \sum_{\underline{y}} \psi_N^m(\underline{y}) \cdot \left(\psi_N^m(\underline{y})^{-\frac{1}{\rho}} \cdot p_N(\underline{y}|\underline{x}^m)^{\frac{1}{\rho}} \cdot \sum_{m' \neq m} \left(\frac{p_N(\underline{y}|\underline{x}^{m'})}{p_N(\underline{y}|\underline{x}^m)} \right)^{\lambda} \right)^{\rho} , \quad \lambda, \rho \geq 0 . \end{aligned} \tag{5}$$

By invoking the Jensen inequality in (5), the DS2 bound results [12]:

$$P_{e|m} \leq \left(\sum_{m' \neq m} \sum_{\underline{y}} p_N(\underline{y}|\underline{x}^m)^{\frac{1}{\rho}} \cdot \psi_N^m(\underline{y})^{1-\frac{1}{\rho}} \cdot \left(\frac{p_N(\underline{y}|\underline{x}^{m'})}{p_N(\underline{y}|\underline{x}^m)} \right)^{\lambda} \right)^{\rho} , \quad 0 \leq \rho \leq 1 , \quad \lambda \geq 0 . \tag{6}$$

Let $G_N^m(\underline{y})$ be an arbitrary non-negative function of $\underline{y}$, and let the probability density function $\psi_N^m(\underline{y})$ be:

$$\psi_N^m(\underline{y}) = \frac{G_N^m(\underline{y}) \cdot p_N(\underline{y}|\underline{x}^m)}{\sum_{\underline{y}} G_N^m(\underline{y}) \cdot p_N(\underline{y}|\underline{x}^m)} . \tag{7}$$

The functions $G_N^m(\underline{y})$ and $\psi_N^m(\underline{y})$ are referred to as the un-normalized and normalized tilting measures respectively. The substitution of (7) into (6) yields:

$$\begin{aligned} P_{e|m} \leq & \left(\sum_{\underline{y}} G_N^m(\underline{y}) \cdot p_N(\underline{y}|\underline{x}^m) \right)^{1-\rho} \\ & \cdot \left\{ \sum_{m' \neq m} \sum_{\underline{y}} p_N(\underline{y}|\underline{x}^m) \cdot G_N^m(\underline{y})^{1-\frac{1}{\rho}} \cdot \left(\frac{p_N(\underline{y}|\underline{x}^{m'})}{p_N(\underline{y}|\underline{x}^m)} \right)^{\lambda} \right\}^{\rho} , \quad 0 \leq \rho \leq 1 , \ \lambda \geq 0 . \end{aligned} \tag{8}$$

The upper bound (8) was also derived in [7, Eq. (62)]. For the class of memoryless channels, and for the choice of $\psi_N^m(\underline{y})$ as $\psi_N^m(\underline{y}) = \prod_{i=1}^{N} \psi^m(y_i)$ (recalling that the function ψ^m may also depend on the transmitted codeword $\underline{x}^m$), the upper bound (8) is relatively easily evaluated (similarly to the standard union bounds) for particular block codes. In that case, (8) is calculable in terms of the distance spectrum, not requiring the fine details of the code structure. Moreover, (8) is also amenable to some generalizations, such as for the class of discrete memoryless channels with arbitrary input and output alphabets.

3. Interconnections and Observations

3.1. The random coding version of the DS2 bound

We show here that the random coding version of the DS2 bound coincides with the well known Gallager 1965 bound for random codes.

For the ensemble of random codes, where the N-length codewords are randomly and independently selected with respect to the input distribution $q_N(\underline{x})$, the DS2 bound yields the following upper bound on the conditional decoding error probability:

$$\begin{aligned} P_{e|m} \quad &\leq \quad (M-1)^{\rho} \cdot \left(\sum_{\underline{y}} p_N(\underline{y}|\underline{x}^m) \cdot G_N^m(\underline{y}) \right)^{1-\rho} \\ &\cdot \quad \left\{ \sum_{\underline{y}} p_N(\underline{y}|\underline{x}^m) \cdot G_N^m(\underline{y})^{1-\frac{1}{\rho}} \cdot \sum_{\underline{x}'} q_N(\underline{x}') \cdot \left(\frac{p_N(\underline{y}|\underline{x}')}{p_N(\underline{y}|\underline{x}^m)} \right)^{\lambda} \right\}^{\rho} , \end{aligned} \tag{9}$$

where $0 \leq \rho \leq 1$ and $\lambda \geq 0$.

The optimal non-negative function $G_N^m(\underline{y})$ minimizing (9) is:

$$G_N^m(\underline{y}) = \left(\sum_{\underline{x}'} q_N(\underline{x}') \cdot \left(\frac{p_N(\underline{y}|\underline{x}')}{p_N(\underline{y}|\underline{x}^m)} \right)^{\lambda} \right)^{\rho} . \tag{10}$$

The substitution of (10) into (9) gives:

$$P_{e|m} \leq (M-1)^{\rho} \cdot \sum_{\underline{y}} \left\{ p_N(\underline{y}|\underline{x}^m) \cdot \left(\sum_{\underline{x}'} q_N(\underline{x}') \cdot \left(\frac{p_N(\underline{y}|\underline{x}')}{p_N(\underline{y}|\underline{x}^m)} \right)^{\lambda} \right)^{\rho} \right\} . \tag{11}$$

After averaging over the transmitted codeword $\underline{x}^m$, which is randomly selected with respect to the input distribution $q_N(\underline{x}^m)$, (11) yields the following upper bound on the ML decoding error probability:

$$P_e \leq (M-1)^{\rho} \cdot \sum_{\underline{y}} \left\{ \left(\sum_{\underline{x}} q_N(\underline{x}) \cdot p_N(\underline{y}|\underline{x})^{1-\lambda\rho} \right) \cdot \left(\sum_{\underline{x}'} q_N(\underline{x}') \cdot p_N(\underline{y}|\underline{x}')^{\lambda} \right)^{\rho} \right\} . \tag{12}$$

Letting $\lambda = \frac{1}{1+\rho}$, the standard random coding Gallager bound [15] results. It is hence demonstrated that in the standard random coding setting, no penalty is incurred for invoking the Jensen inequality in the optimized DS2 bound.

3.2. Relations to the Gallager-Fano 1961 bound

In this subsection we present the Gallager-Fano 1961 bound [17] and study its relation to the DS2 bound.

3.2.1. The 1961 Gallager-Fano bound

In his monograph on low-density parity-check codes [17], Gallager introduced a general upper bound on the ML decoding error probability for block codes operating over binary-input and output-symmetric channels. This bound depends on the distance spectrum of the N-length block code (or ensemble of codes), and therefore it can be applied to both ensembles as well as specific codes (as opposed to the Gallager 1965 bound which applies essentially to the ensemble of random block codes). The derivation of this bound is based on Fano's 1961 bounding technique [14] for random codes, which is adapted in [17] to specific codes. The concept of this bound is shortly detailed in the following:
Let $\underline{x}^m$ be the transmitted codeword and define the tilted ML metric:

$$D_m(\underline{x}^{m'}, \underline{y}) = \ln \left(\frac{f_N^m(\underline{y})}{p_N(\underline{y}|\underline{x}^{m'})} \right) , \qquad (13)$$

where $\underline{x}^{m'}$ is an arbitrary codeword, $p_N(\underline{y}|\underline{x})$ is the conditional transition probability of the channel, and $f_N^m(\underline{y})$ is an arbitrary function which is positive if $p_N(\underline{y}|\underline{x}^{m'})$ is positive for any m' (and may also depend on the transmitted message). Note that $D_m(\cdot, \cdot)$ is in general not computable at the receiver; it is used here as a conceptual tool to evaluate the upper bound on the decoding error probability. If ML decoding is applied, an error occurs if for some $m' \neq m$:

$$D_m(\underline{x}^{m'}, \underline{y}) \leq D_m(\underline{x}^m, \underline{y}) . \qquad (14)$$

The received set Y^N is divided into two **disjoint** subsets:

$$\begin{aligned} Y^N &= Y_g^N \bigcup Y_b^N , \\ Y_g^N &= \left\{ \underline{y} : D_m(\underline{x}^m, \underline{y}) \leq Nd \right\} , \\ Y_b^N &= \left\{ \underline{y} : D_m(\underline{x}^m, \underline{y}) > Nd \right\} , \end{aligned} \qquad (15)$$

where d is an arbitrary real number.

The conditional ML decoding error probability can be expressed as a sum of two terms:

$$P_{e|m} = \text{Prob}\,(\text{error}\,, \underline{y} \in Y_b^N) + \text{Prob}\,(\text{error}\,, \underline{y} \in Y_g^N) , \qquad (16)$$

which leads to the following upper bound on the conditional decoding error probability:

$$P_{e|m} \leq \text{Prob}\,(\underline{y} \in Y_b^N) + \text{Prob}\,(\text{error}\ , \underline{y} \in Y_g^N)\ . \tag{17}$$

Inequality (17) is the starting point of many efficient bounds, for example the Berlekamp tangential bound [4], Hughes bound, Poltyrev tangential sphere bound [24] and Engdahl and Zigangirov bound [13]. In the Fano-Gallager approach, the regions Y_g^N, Y_b^N are related to the choice of the arbitrary function f in (13). At this stage, Fano [14] proceeded with the Chernoff bounding technique and the random coding approach, while Gallager bound [17] is better suited to treat particular codes.

Based on the Chernoff bound:

$$\text{Prob}\,(\underline{y} \in Y_b^N) \leq E(e^{sW})\ , \quad s \geq 0\ , \tag{18}$$

where

$$W = \ln\left(\frac{f_N^m(\underline{y})}{p_N(\underline{y}|\underline{x}^m)}\right) - Nd\ . \tag{19}$$

The second term in (17) is also upper bounded by a combination of the union and the Chernoff bounds, which then yields:

$$\begin{aligned}
&\text{Prob}\,(\text{error}\ , \underline{y} \in Y_g^N) \\
&= \text{Prob}\,\Big(D_m(\underline{x}^{m'}, \underline{y}) \leq D_m(\underline{x}^m, \underline{y}) \text{ for some } m' \neq m\,, \quad \underline{y} \in Y_g^N\Big) \\
&\leq \sum_{m' \neq m} \text{Prob}\,\Big(D_m(\underline{x}^{m'}, \underline{y}) \leq D_m(\underline{x}^m, \underline{y})\,,\ D_m(\underline{x}^m, \underline{y}) \leq Nd\Big) \\
&\leq \sum_{m' \neq m} E\Big(\exp(t \cdot Z_{m'} + r \cdot W)\Big)\ , \quad t \geq 0,\ r \leq 0\ ,
\end{aligned} \tag{20}$$

where, based on (13), $Z_{m'} = D_m(\underline{x}^m, \underline{y}) - D_m(\underline{x}^{m'}, \underline{y}) = \ln\left(\frac{p_N(\underline{y}|\underline{x}^{m'})}{p_N(\underline{y}|\underline{x}^m)}\right)$ and W is defined in (19). In [17], Gallager assumed a binary linear block code operating over a binary-input, output-symmetric and memoryless channel. Under these assumptions, the discussion in [17] was restricted to functions $f_N^m(\underline{y})$ which can be expressed in the product form:

$$f_N^m(\underline{y}) = \prod_{i=1}^{N} f(y_i)\ , \tag{21}$$

(where the function f does not depend here on the index i). For simplifying the derivation of the Gallager-Fano 1961 bound [17], it was also assumed that the non-negative function f is **even**, i.e: $f(y) = f(-y)$ for all values of y.

Let $\{S_l\}_{l=0}^N$ designate the distance spectrum of the considered block code. Based on (17)-(20) and by optimally setting $t = \frac{1-r}{2}$, the Gallager-Fano 1961 bound [17] reads:

$$P_e \leq g(s)^N \cdot \exp(-Nsd) + \sum_{l=0}^{N} S_l \cdot [h(r)]^l \cdot [g(r)]^{N-l} \cdot \exp(-Nrd) \; , \tag{22}$$

where $s \geq 0$, $r \leq 0$, $-\infty < d < +\infty$, $t \geq 0$. By the symmetry of the channel and the function f:

$$\begin{aligned} g(s) &= \frac{1}{2} \cdot \sum_y \left([p_0(y)]^{1-s} + [p_1(y)]^{1-s} \right) \cdot f(y)^s \; , \\ h(r) &= \sum_y [p_0(y) \cdot p_1(y)]^{\frac{1-r}{2}} \cdot f(y)^r \; . \end{aligned} \tag{23}$$

The upper bound (22) on the ML decoding error probability depends on the distance spectrum of the considered block code, and therefore it applies to fixed codes and structured ensembles of codes. It can be also verified (though not mentioned in [17]) that the optimization of the parameter d in (22), yielding the tightest bound within this family, assumes the expression:

$$P_e \leq 2^{H(\rho)} \cdot g(s)^{N(1-\rho)} \cdot \left(\sum_{l=0}^{N} S_l \cdot h(r)^l \cdot g(r)^{N-l} \right)^{\rho} \; . \tag{24}$$

Here $\rho = \frac{s}{s-r}$ $(0 \leq \rho \leq 1)$ and $H(\rho) = -\rho \log_2(\rho) - (1-\rho)\log_2(1-\rho)$ is the binary entropy function. The optimization of the function f in (22) yields an implicit solution (as indicated in [17, Eq. (3.40)]). The following sub-optimal and explicit solution for the non-negative and symmetric function f was proposed by Gallager [17, Eq. (3.41)]:

$$f(y) = k \cdot \left\{ \frac{\left[P_0(y)^{\frac{1-r}{2}} + P_1(y)^{\frac{1-r}{2}} \right]^2}{P_0(y)^{1-s} + P_1(y)^{1-s}} \right\}^{\frac{1}{s-r}} \; , \tag{25}$$

where the constant k in Eq. (25) is arbitrary and cancels out in the bound.

3.2.2. The connection between the Gallager-Fano 1961 bound and the DS2 bound

The optimization over the parameter d (or e^{-Nd}) in the Gallager-Fano 1961 bound on the conditional decoding error probability (see Eq. (17)-(20)) and the setting $\rho = \frac{s}{s-r}$ $(0 \leq \rho \leq 1$ as $s \geq 0$

and $r \leq 0$) gives:

$$P_{e|m} \leq 2^{H(\rho)} \cdot \left(\sum_{\underline{y}} p_N(\underline{y}|\underline{x}^m) \cdot \left(\frac{f_N^m(\underline{y})}{p_N(\underline{y}|\underline{x}^m)} \right)^s \right)^{1-\rho} \cdot \left(\sum_{m' \neq m} \sum_{\underline{y}} p_N(\underline{y}|\underline{x}^m) \cdot \left(\frac{p_N(\underline{y}|\underline{x}^{m'})}{p_N(\underline{y}|\underline{x}^m)} \right)^t \cdot \left(\frac{f_N^m(\underline{y})}{p_N^m(\underline{y}|\underline{x}^m)} \right)^{s\left(1-\frac{1}{\rho}\right)} \right)^{\rho} , \quad 0 \leq \rho \leq 1 . \tag{26}$$

as was first indicated by Divsalar [7, Eqs. (71), (72)].

Divsalar [7] has renamed t by λ (where t is the non-negative parameter introduced in (20)) and has also set:

$$G_N^m(\underline{y}) = \left(\frac{f_N^m(\underline{y})}{p_N(\underline{y}|\underline{x}^m)} \right)^s , \tag{27}$$

deriving then the DS2 bound (8) with an additional factor of $2^{H(\rho)}$ (where $1 \leq 2^{H(\rho)} \leq 2$). This demonstrates the superiority of the DS2 bound over the Gallager-Fano 1961 bounding technique when applied to a particular code or ensemble of codes. It has been demonstrated in [33] that the Gallager-Fano 1961 bound equals the 1965 Gallager random coding bound (12) up to the $2^{H(\rho)}$ coefficient, where as shown before, the latter agrees with the optimized DS2 bound.

3.3. A geometric interpretation of the Gallager type bounds

The connections between the Gallager-Fano tilting measure and the Duman and Salehi normalized and un-normalized tilting measures (which are designated here by $f_N^m(\underline{y})$, $\psi_N^m(\underline{y})$ and $G_N^m(\underline{y})$ respectively) are indicated in Eqs. (7), (27), and they also provide some geometric interpretations of various reported bounds. The measure $f_N^m(\underline{y})$ in the Gallager-Fano 1961 setting, which in general does not imply a product form, entails a geometric interpretation associated with the conditions in the inequalities (15), specifying the disjoint regions Y_g^N, $Y_b^N \subseteq Y^N$. The geometric interpretation of the Gallager-Fano 1961 bound is not necessarily unique, as measures $f_N^m(\underline{y})$ of different functional structure may imply equivalent conditions in the inequality:

$$\ln \left(\frac{f_N^m(\underline{y})}{p_N(\underline{y}|\underline{x}^m)} \right) \leq Nd . \tag{28}$$

This non-uniqueness property is due to the shifting and factoring invariance of inequality (28) and since the parameter d is also subjected to optimization. We demonstrate here the connection between the non-unique measure $f_N^m(\underline{y})$ and its associated decision region, and to that end we discuss some reported upper bounds on the ML decoding error probability for the binary-input

AWGN channel. All of these upper bounds are based on inequality (16). In order to demonstrate the non-uniqueness feature of the Gallager-Fano tilting measure, we shall focus on the Divsalar bound [7].

The conditional density function for the binary-input AWGN channel is:

$$p_N(\underline{y}|\underline{x}^m) = \prod_{l=1}^{N} \left\{ \frac{1}{\sqrt{2\pi}} \cdot \exp\left[-\frac{1}{2} \cdot (y_l - \gamma \cdot x_l^m)^2 \right] \right\} , \tag{29}$$

where $\gamma = \sqrt{\frac{2RE_b}{N_0}}$. Here x_l^m designates the l-th symbol of the m-th codeword.

Divsalar bound [7] is specified by the associated decision region

$$Y_g^N = \left\{ \underline{y} \,|\, \sum_{l=1}^{N} (y_l - \eta\gamma \cdot x_l^m)^2 \leq N \cdot r^2 \right\} , \tag{30}$$

which implies that the region Y_g^N is an N-dimensional sphere with radius $\sqrt{N \cdot r^2}$ and a center that is located along the line connecting the origin to the codeword $\underline{x}^m$ and at distance $\sqrt{N}\,\eta\gamma$ from the origin (the coordinates of the codeword $\underline{x}^m$ are either 1 or -1). The parameters r, η are analytically optimized in Divsalar bound [7]. It can be verified that the following Fano-Gallager tilting measures in (28) can be associated with the same decision region (30):

$$f_N^m(\underline{y}) = \prod_{l=1}^{N} \left\{ \exp\left(\frac{1}{2} \left(\frac{1}{\eta} - 1 \right) \cdot (y_l)^2 \right) \right\} , \tag{31a}$$

$$f_N^m(\underline{y}) = \prod_{l=1}^{N} \left\{ \exp\left(\frac{\gamma}{2} \cdot (1 - \eta) \cdot x_l^m \cdot y_l \right) \right\} , \tag{31b}$$

$$f_N^m(\underline{y}) = \prod_{l=1}^{N} \left\{ \exp\Big(-\theta \cdot (y_l - \phi \cdot x_l^m)^2 \Big) \right\} , \text{where } \phi = \gamma \cdot \left(\eta + \frac{1-\eta}{2\theta} \right) , \ \theta \neq 0 , \tag{31c}$$

as a shift in the value of the parameter d in (15) before its optimization has no implications on the bound. Although the three tilting measures $f_N^m(\underline{y})$ imply the same decision region Y_g^N, the second and third measures ((31b),(31c)) are amenable to generalizations for the class of fully interleaved fading channels, as opposed to the first measure which demonstrates a different functional behavior (as the exponential term is quadratic in $\{y_i\}$).

4. Special Cases of the DS2 Bound

In this section we demonstrate that many reported bounds can be considered as special cases of the generalized DS2 bound. [1] Some of these observations have been reported in [7],[33].

4.1. Shulman and Feder bound

We derive here the Shulman and Feder bound [36] as a particular case of the DS2 bound. Let C be a binary and linear block code of length N and rate $R = \frac{\log_2 M}{N}$ bits per channel use, with a distance spectrum $\{S_k\}_{k=0}^{N}$. Suppose that the all-zero codeword ($\underline{x}^0$) is transmitted over a memoryless binary-input and symmetric-output channel with a transition probability $p(y|x)$ and that ML decoding is performed. In order to be able to express the upper bound (8) in terms of the distance spectrum of the block code C, we restrict our attention here to the special case where the un-normalized Duman and Salehi tilting measure $G_N^0(\underline{y})$ can be expressed in a product form: $G_N^0(\underline{y}) = \prod_{i=1}^{N} g(y_i)$. Based on the upper bound (8):

$$P_{e|0} \leq \left(\sum_y g(y) \cdot p(y|0)\right)^{N(1-\rho)} \cdot 2^{-N(1-R)\rho}$$

$$\left\{\sum_{l=0}^{N} \left(\frac{S_l}{2^{-N(1-R)}\binom{N}{l}}\right) \binom{N}{l} \left(\sum_y p(y|0) g(y)^{1-\frac{1}{\rho}}\right)^{N-l} \left(\sum_y p(y|0)^{1-\lambda} p(y|1)^{\lambda} g(y)^{1-\frac{1}{\rho}}\right)^{l}\right\}^{\rho}$$

$$\leq \left(\sum_y g(y) \cdot p(y|0)\right)^{N(1-\rho)} \cdot \left(\max_{0 \leq l \leq N} \frac{S_l}{2^{-N(1-R)}\binom{N}{l}}\right)^{\rho} \cdot 2^{-N(1-R)\rho}$$

$$\cdot \left(\sum_y p(y|0) \cdot g(y)^{1-\frac{1}{\rho}} + \sum_y p(y|0)^{1-\lambda} \cdot p(y|1)^{\lambda} \cdot g(y)^{1-\frac{1}{\rho}}\right)^{N\rho} . \tag{32}$$

By setting:

$$g(y) = \left[\frac{1}{2} \cdot p(y|0)^{\frac{1}{1+\rho}} + \frac{1}{2} \cdot p(y|1)^{\frac{1}{1+\rho}}\right]^{\rho} \cdot p(y|0)^{\frac{-\rho}{1+\rho}} , \; \lambda = \frac{1}{1+\rho} , \tag{33}$$

[1]The first version of the Duman and Salehi bounds which is introduced in this section is also demonstrated to be a particular case of the DS2 bound. Therefore, we introduced the latter bound in Section 2, before the presentation of their first version bound in Section 4.

and after some straight forward algebra invoking the channel's symmetry, we get:

$$P_{e|0} \leq 2^{NR\rho} \cdot \left(\sum_y \left[\frac{1}{2}\, p(y|0)^{\frac{1}{1+\rho}} + \frac{1}{2}\, p(y|1)^{\frac{1}{1+\rho}} \right]^{1+\rho} \right)^N \cdot \left(\max_{0 \leq l \leq N} \frac{S_l}{2^{-N(1-R)} \binom{N}{l}} \right)^\rho , \tag{34}$$

which agrees with Shulman and Feder bound [36]. Based on the DS2 bound, a generalization of the Shulman and Feder bound is detailed in Appendix A.

4.2. Upper bounds on the ML decoding error probability for the binary-input AWGN channel

In this sub-section, we present various upper bounds on the ML decoding error probability for the binary-input AWGN channels. We demonstrate here that these reported bounds, which were originally derived independently, are special cases of the DS2 bound. In general, this is done by choosing for every such upper bound an appropriate probability tilting measure, and then by re-deriving this bound as a special case of the DS2 bound of Section 2.

4.2.1. Duman and Salehi bound (first version)

The first version of the Duman and Salehi bound [11] for a binary-input AWGN channel is a special case of the DS2 bound [12], where the normalized tilting measure is:

$$\psi_N^m(\underline{y}) = \prod_{l=1}^N \psi^m(y_l) = \prod_{l=1}^N \left\{ \sqrt{\frac{\alpha}{2\pi}} \cdot \exp\left[-\frac{\alpha}{2} \cdot \left(y_l - \frac{\beta}{\alpha} \cdot \sqrt{\frac{2E_s}{N_0}} \cdot x_l^m \right)^2 \right] \right\} . \tag{35}$$

We assume here that the components of the codeword $\underline{x}^m$ are either -1 or $+1$ for a '0' or '1' input respectively. Based on the connection between the normalized and the un-normalized tilting measures (7), and since the upper bound (8) is invariant to scaling of the un-normalized tilting, it is also a particular case of (8), where the un-normalized tilting measure is:

$$G_N^m(\underline{y}) = \prod_{l=1}^N \left\{ \exp\left[-\frac{\alpha}{2} \cdot \left(y_l - \frac{\beta}{\alpha} \cdot \sqrt{\frac{2E_s}{N_0}} \cdot x_l^m \right)^2 + \frac{1}{2} \cdot \left(y_l - \sqrt{\frac{2E_s}{N_0}} \cdot x_l^m \right)^2 \right] \right\} . \tag{36}$$

A binary and linear block code is then partitioned to constant Hamming weight subcodes (each of them also includes the all-zero codeword). The conditional ML decoding error probability is upper bounded by (6), where the all-zero vector is assumed to be the transmitted codeword. Due to the channel symmetry and the linearity of the code, an overall union bound over the subcodes yields:

$$P_e = P_{e|0} \leq \sum_{d=d_{\min}}^N P_{e|0}(d) , \tag{37}$$

where $d_{\min}$ is the minimum Hamming distance of the N-length block code and $P_{e|0}(d)$ is the conditional decoding error probability with respect to the subcode with a constant Hamming weight d. By (6), $P_{e|0}(d)$ is upper bounded,

$$P_{e|0}(d) \leq (S_d)^\rho \cdot \left(\int_{-\infty}^{+\infty} p_0(y)^{\frac{1}{\rho}} \cdot \psi(y)^{1-\frac{1}{\rho}} \, dy \right)^{(N-d)\rho} \cdot \left(\int_{-\infty}^{+\infty} p_0(y)^{\frac{1-\lambda\rho}{\rho}} \cdot p_1(y)^\lambda \cdot \psi(y)^{1-\frac{1}{\rho}} \, dy \right)^{d\rho}, \tag{38}$$

where $0 \leq \rho \leq 1$, $\lambda > 0$, and

$$p_0(y) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left[-\frac{1}{2} \cdot \left(y + \sqrt{\frac{2RE_b}{N_0}} \right)^2 \right], \tag{39}$$

$$p_1(y) = p_0(-y), \quad -\infty < y < +\infty$$

designate the transition probabilities of the binary-input AWGN channel.

By substituting (35),(39) into (38), the optimization over the parameters $\lambda \geq 0$ and β in [11] results:

$$\beta^* = \frac{1 - \frac{d}{N}}{\frac{1}{\alpha} - \frac{d}{N} \cdot (1-\rho)}, \quad \lambda^* = \frac{1}{2} \cdot \left(\beta^* + \frac{1-\beta^*}{\rho} \right). \tag{40}$$

The substitution of (35), (39) and (40) into (38) then yields:

$$P_{e|0}(d) \leq (S_d)^\rho \cdot \alpha^{\frac{N(1-\rho)}{2}} \cdot \left(\alpha - \frac{\alpha-1}{\rho} \right)^{-\frac{N\rho}{2}}$$
$$\cdot \exp\left\{ N \cdot \left(\frac{RE_b}{N_0} \right) \cdot \left[-1 + \frac{(\beta^*)^2 \cdot (1-\rho)}{\alpha} + \frac{\rho \cdot \left(1 - \frac{d}{N}\right) \cdot \left(\beta^* + \frac{1-\beta^*}{\rho} \right)^2}{\alpha - \frac{\alpha-1}{\rho}} \right] \right\}. \tag{41}$$

The bound (41) is the first version of the Duman and Salehi bounds [11] and it is numerically optimized over the range: $0 \leq \alpha < \frac{1}{1-\rho}$, $0 < \rho \leq 1$.

4.2.2. Viterbi and Viterbi bound (first version)

The Viterbi and Viterbi bound [39] is an upper bound on the ML decoding error probability for BPSK modulated block codes operating over a binary-input AWGN channel. It can be verified that it is also a particular case of the first version of Duman and Salehi bounds by substituting $\alpha = 1$ in (40) and (41), which yields:

$$P_e \leq (S_d)^\rho \cdot \exp\left(\frac{-NRE_b}{N_0} \cdot \frac{\left(\frac{d}{N}\right) \cdot \rho}{1 - \frac{d}{N} \cdot (1-\rho)} \right), \quad 0 \leq \rho \leq 1\,. \tag{42}$$

The optimization over the parameter ρ gives then the Viterbi & Viterbi upper bound [39], which reads:

$$P_e \leq \exp\Big(-N \cdot E_{v_1}(r_d)\Big) \ , \tag{43}$$

where

$$r_d = \frac{\ln(S_d)}{N} \ , \ c = \frac{E_s}{N_0} = \frac{RE_b}{N_0} \ , \tag{44}$$

$$E_{v_1}(r_d) = \begin{cases} \left(\frac{d}{N}\right) c - r_d & , \quad 0 \leq \frac{r_d}{c} \leq \frac{d}{N} \cdot \left(1 - \frac{d}{N}\right) \ , \\ \left(\sqrt{c} - \sqrt{\frac{r_d \cdot \left(1 - \frac{d}{N}\right)}{\frac{d}{N}}}\right)^2 & , \quad \frac{d}{N} \cdot \left(1 - \frac{d}{N}\right) \leq \frac{r_d}{c} \leq \frac{\frac{d}{N}}{1 - \frac{d}{N}} \ . \end{cases} \tag{45}$$

4.2.3. Viterbi and Viterbi bound (second version)

The second version of the Viterbi and Viterbi bounds [38] is based on the Gallager-Fano 1961 bound and it reads:

$$P_{e|0}(l) \leq \exp\Big(-N \cdot E_{v_2}(r_l)\Big) \ , \tag{46}$$

where $r_l = \frac{\ln(S_l)}{N}$ and,

$$E_{v_2}(r_l) = \max_{0 \leq \rho \leq 1} \left\{ -\rho r_l + \frac{l}{N} \cdot \ln\Big(\bar{h}(\rho)\Big) + \left(1 - \frac{l}{N}\right) \cdot \ln\Big(\bar{g}(\rho)\Big) - (1-\rho) \cdot \ln\Big(\bar{h}(\rho) + \bar{g}(\rho)\Big) \right\} \ . \tag{47}$$

Based on the symmetry of the binary-input and memoryless channel, we get from (23),(25) and the substitution $\rho = \frac{s}{s-r}$ (where $s \geq 0$, $r \leq 0$ and $0 \leq \rho \leq 1$):

$$\begin{aligned} \bar{h}(\rho) &\triangleq h(r) = \sum_y \Big[p_0(y)^{\frac{1}{1+\rho}} + p_0(-y)^{\frac{1}{1+\rho}}\Big]^{-(1-\rho)} \cdot \Big[p_0(y) \cdot p_0(-y)\Big]^{\frac{1}{1+\rho}} \ , \\ \bar{g}(\rho) &\triangleq g(r) = \sum_y \Big[p_0(y)^{\frac{1}{1+\rho}} + p_0(-y)^{\frac{1}{1+\rho}}\Big]^{-(1-\rho)} \cdot p_0(y)^{\frac{2}{1+\rho}} \ . \end{aligned} \tag{48}$$

For continuous output channels, the summation in (48) should of course be replaced by integrals. This bound is again a special case of the DS2 bound, as noticed by substituting in (8) the un-normalized tilting measure:

$$G_N^m(\underline{y}) = \prod_{l=1}^{N} \left\{ \left(p(y_l|0)^{\frac{1}{1+\rho}} + p(y_l|1)^{\frac{1}{1+\rho}} \right)^{\rho} \cdot p(y_l|0)^{-\frac{\rho}{1+\rho}} \right\} . \tag{49}$$

4.2.4. Divsalar bound

Divsalar bound [7] is in fact a particular case of the Gallager-Fano 1961 bounding technique for a binary-input AWGN channel (see subsection 3.3). This bound which is given in [7] in closed form reads:

$$P_e \leq \sum_{d=d_{\min}}^{N} \min \left\{ \exp\Big(-N \cdot E(c,d,\beta^*)\Big) \, , \; S_d \cdot Q(\sqrt{2dc}) \right\} \, , \tag{50}$$

where

$$E(c,d,\beta) = -r_d + \frac{1}{2} \cdot \ln \Big[\beta + (1-\beta) \cdot \exp(2 \cdot r_d) \Big] + \frac{\left(\frac{d}{N}\right) \cdot \beta}{1 - \left(\frac{d}{N}\right) \cdot (1-\beta)} \, , \tag{51}$$

$$\beta^* = \sqrt{ \left(\frac{1-\frac{d}{N}}{\frac{d}{N}} \right) \frac{2c}{1-\exp(-2\, r_d)} + \left(\frac{1-\frac{d}{N}}{\frac{d}{N}} \right)^2 \Big[(1+c)^2 - 1 \Big] } - \left(\frac{1-\frac{d}{N}}{\frac{d}{N}} \right) (1+c), \tag{52}$$

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{+\infty} \exp\left(-\frac{t^2}{2} \right) \, dt \, , \tag{53}$$

and the parameters c, r_d are specified in (44). The bound is valid for $\frac{E_b}{N_0}$ values which are not below a certain threshold:

$$\frac{E_b}{N_0} \geq \frac{1}{R} \cdot \max_{0 \leq d \leq N(1-R)} \left(\frac{(1-\exp(-2\, r_d))\ (1-\frac{d}{N})}{2 \cdot \left(\frac{d}{N}\right)} \right) . \tag{54}$$

Based on (7) and (27), one obtains the following relation between the Fano-Gallager tilting measure ($f_N^m(\underline{y})$) and the Duman and Salehi normalized tilting measure ($\psi_N^m(\underline{y})$):

$$\psi_N^m(\underline{y}) = \frac{f_N^m(\underline{y})^s \cdot p_N^m(\underline{y}|\underline{x}^m)^{1-s}}{\sum_{\underline{y}} f_N^m(\underline{y})^s \cdot p_N^m(\underline{y}|\underline{x}^m)^{1-s}} \, . \tag{55}$$

By substituting into (55) the Fano-Gallager tilting measure (f_N^m) in (31c) and the conditional probability density function (p_N^m) for the binary-input AWGN channel (29), then the normalized tilting measure (ψ_N^m) for Divsalar bound coincides with the normalized tilting measure in (35). This result is verified by setting the parameters as follows:

$$\alpha = (2\theta - 1)s + 1\,, \quad \beta = (2\theta - 1)\eta s + 1\,. \tag{56}$$

This is the grounds for the observation in [7] that Divsalar bound is a closed form version of the Duman and Salehi (first version) bound [11].

4.2.5. Engdahl and Zigangirov bound

In Engdahl and Zigangirov bound [13] which was derived for a binary-input AWGN channel, the decision region Y_g^N associated with the transmitted codeword $\underline{x}^m$ is an N-dimensional plane:

$$Y_g^N = \left\{ \underline{y} \,|\, \sum_{l=1}^{N} y_l \cdot x_l^m \geq Nd \right\}\,. \tag{57}$$

The motivation for (57) is that inside the good region (i.e, $\underline{y} \in Y_g^N$), the correlation between the received vector $\underline{y}$ and the transmitted codeword $\underline{x}^m$ is not expected to be too small. Therefore, it is above a certain threshold (Nd), where the parameter d is to be optimized, as to get the tightest upper bound within the family of bounds (28) and the associated decision region Y_g^N in (57). The following Fano-Gallager tilting measure can be associated with the same decision region Y_g^N:

$$f_N^m(\underline{y}) = \prod_{l=1}^{N} \left\{ \exp\left(-\frac{1}{2} \cdot y_l^2 + \beta \cdot y_l \cdot x_l^m \right) \right\}\,, \tag{58}$$

where $\beta \neq \gamma$ (γ is introduced in (29)). The free parameter β in (58) demonstrates again that there might be some functionally different Fano-Gallager tilting measures which imply the same decision region (based on the definition of the regions in (15)). It is interesting to note, that for this specific choice of a decision region (57), it was demonstrated in [13] that there is no need to invoke the Chernoff bounds for the binary-input AWGN channel, and the two terms in the right hand side of (17) can be exactly calculated.

4.2.6. The Chernoff version of various bounds for the binary-input AWGN channel

In his paper [7], Divsalar derived simplified Chernoff versions of some upper bounds, which result as particular cases of the Fano-Gallager 1961 bounding technique. These simplified Chernoff versions include the following bounds:

The Chernoff version of the sphere bound: This bound results as a particular case of the decision region in Divsalar bound (31), where $\eta = 1$.

The Chernoff version of the tangential bound [4]: This version coincides with the first version of the Viterbi and Viterbi bound [39]. It can be shown that the relevant Fano-Gallager decision region is a plane, which is also the case for Engdahl and Zigangirov bound [13]. However, as noted above, in the latter bound the Chernoff bounding technique is not invoked, improving thus the bound for a finite block length.

The Chernoff version of the tangential sphere bound: The tangential sphere bound of Poltyrev [24] happens often to be the tightest reported upper bound for block codes which are transmitted over a binary-input AWGN channel and ML decoded (see e.g., [30]–[32]). However, in the random coding setting, it fails to reproduce the random coding exponent [24] while the DS2 bound does. This bound involves a numerical solution of an associated optimization equation ([30, Eq. (6)]), and it is therefore not expressed in closed form. In his paper [7], Divsalar derived a simplified Chernoff version of the tangential sphere bound. That upper bound is shown in [7] to have the same error exponent as the tangential sphere bound of Poltyrev, and therefore the loosening of the tangential sphere bound by invoking the Chernoff bounding technique, does not carry any implications on the tightness of the bound for asymptotically infinite block length.

Due to the connection of the Fano-Gallager tilting measure to the Duman and Salehi variation, it is evident that the Chernoff versions of these bounds can be also viewed as special cases of the DS2 bound. It should be emphasized that the mentioned bounds above were originally developed without resorting to Chernoff bounding technique, yielding thus tighter versions of these bounds.

4.3. Upper bounds on the ML decoding error probability for fully interleaved fading channels with perfect channel state information at the receiver

We demonstrate here various variations of the DS2 bound when applied to the class of fully interleaved Rayleigh fading channels with perfect channel state information at the receiver. This problem is treated in detail in [8],[34],[26]-[28] and here we present in a comparative fashion some insightful results. The model is:

$$y = a \cdot x + n \ , \tag{59}$$

where y stands for the received signal, x stands for the BPSK modulated input signal (that is $\pm\sqrt{2E_s}$) and n designates the additive zero mean and $\frac{N_0}{2}$ variance Gaussian noise component. The fading a is assumed to be perfectly known at the receiver and hence is considered to be real valued, as the receiver compensates for any phase rotation. Due to the ideal interleaving, the channel is assumed to be memoryless. The bounds are based on first decomposing the code to constant-weight subcodes (where every subcode also includes the all-zero codeword), over which a union bound is invoked as in (37).

4.3.1. Optimized DS2 bound

In [26] the measure:

$$\Psi(\underline{y},\, \underline{a}) = \prod_{l=1}^{N} \psi(y_l,\, a_l)\ , \tag{60}$$

is optimized to yield the best possible DS2 bound (6), where (y, a) are interpreted as the available measurements at the receiver. In [26],[27],[28] the results are extended to the Rician channel.

4.3.2. Exponential tilting measure

In [34], a sub-optimal selection for ψ in (60) is suggested which in fact is motivated by the Duman and Salehi (first version) bound [11]. An exponential tilting is also applied in [34] to the fading sample a (treated as a measurement), which gives rise to the exponential tilting measure:

$$\psi(y, a) = \frac{\sqrt{\frac{\alpha}{2\pi}} \cdot \exp\left[-\frac{\alpha}{2} \cdot \left(y - au \cdot \sqrt{\frac{2E_s}{N_0}}\right)^2 - \frac{\alpha v^2 a^2 E_s}{N_0}\right] \cdot p(a)}{\displaystyle\int_0^{+\infty} p(a) \cdot \exp\left(-\frac{\alpha v^2 a^2 E_s}{N_0}\right) \cdot da}\ , \tag{61}$$

where $\alpha \geq 0$, $-\infty < u < +\infty$, $-\infty < v < +\infty$, and $p(a)$ designates the probability density function of the independent fading samples. That yields a closed form upper bound in [34], which reads:

$$P_{e|0}(d) \leq (S_d)^{\rho} \cdot \alpha^{\frac{-N(1-\rho)}{2}} \cdot \left(\alpha - \frac{\alpha - 1}{\rho}\right)^{-\frac{N\rho}{2}} \cdot \left(\frac{1}{1+t}\right)^{N(1-\rho)} \cdot \left(\frac{1}{1+\varepsilon}\right)^{d\rho} \cdot \left(\frac{1}{1+v}\right)^{(N-d)\rho}\ , \tag{62}$$

with:

$$t = \frac{\alpha \cdot \nu^2 \cdot E_s}{N_0} \, , \tag{63a}$$

$$v = \frac{E_s}{N_0} \cdot \left[\alpha(u^2 + \nu^2) - \frac{\alpha(u^2 + \nu^2) - 1}{\rho} - \frac{\left(\alpha u - \frac{\alpha u - 1}{\rho} \right)^2}{\alpha - \frac{\alpha - 1}{\rho}} \right] \, , \tag{63b}$$

$$\varepsilon = \frac{E_s}{N_0} \cdot \left[\alpha(u^2 + \nu^2) - \frac{\alpha(u^2 + \nu^2) - 1}{\rho} - \frac{\left(\alpha u - \frac{\alpha u - 1}{\rho} - 2\lambda \right)^2}{\alpha - \frac{\alpha - 1}{\rho}} \right] \, . \tag{63c}$$

This bound is in fact equivalent to the Divsalar and Biglieri bound [8] which has been derived via a geometric extension of the associated decision region in Divsalar bound [7] (by rotating the displaced sphere region in [7]). The bound (61)-(63) also yields to an extension of the Viterbi and Viterbi (first version) bound for fully interleaved fading channels and perfect CSI at the receiver (see [34]), by setting:

$$\alpha = 1 \, , \quad u = 1 - \xi\rho \, , \quad \nu = \sqrt{1 - (1 - \xi\rho)^2} \, , \quad \lambda = \frac{1 + \xi \cdot (1 - \rho)}{2} \, . \tag{64}$$

In [34], the DS2 bound associated with the exponential tilting measure (61) is also applied to the fully interleaved Rician fading channel. As a special case, it yields the Duman and Salehi (first version) bound for a binary-input AWGN channel, where the Rician parameter K standing for the power ratio of the direct and the diffused received paths, goes to infinity. The bounds are depicted in Fig. 2 for the 'repeat and accumulate' (RA) codes (an ensemble of turbo-like codes which was introduced by Divsalar, Jin and McEliece [9]) operating over a fully interleaved Rayleigh fading channel. The penalty for the gradual specialization of the DS2 bound by constraining the selection ψ in (60) is explicitly indicated.

5. Gallager-type Bounds for the Mismatched Decoding Regime

In this section we generalize the DS2 bound for a mismatched decoding metric. The basic setting is as before, that is a codeword $\underline{x}$ is conveyed via a channel of transition probability $p_N(\underline{y}|\underline{x})$. The decoder operates in a ML fashion, but may use a mismatched metric $Q_N(\underline{y}|\underline{x})$. Namely, code $\underline{x}^j$ out of the M possible equiprobable codewords is declared if

$$Q_N(\underline{y}|\underline{x}^j) > Q_N(\underline{y}|\underline{x}^i) \, , \quad \forall \, i \neq j \, , \quad i = 1, 2, \ldots, M \, , \tag{65}$$

and ties are randomly resolved. The matched case results when $Q_N(\underline{y}|\underline{x}) = p_N(\underline{y}|\underline{x})$. In general, $Q_N(\underline{y}|\underline{x})$ is not necessarily normalized to yield a probability measure.

We first employ the Duman and Salehi bounding technique in this setting and then examine the performance of random ensembles of codes.

5.1. The mismatched Duman and Salehi bound

The standard Gallager upper bound on the conditional decoding error probability for the mismatched case $P_{e|m}$ (conditioned on the transmitted codeword $\underline{x}^m$) is given by

$$P_{e|m} \leq \sum_{\underline{y}} p_N(\underline{y}|\underline{x}^m) \cdot \left(\sum_{m' \neq m} \left(\frac{Q_N(\underline{y}|\underline{x}^{m'})}{Q_N(\underline{y}|\underline{x}^m)} \right)^{\lambda} \right)^{\rho} , \quad \lambda, \rho \geq 0 \,. \tag{66}$$

By invoking the Duman and Salehi bounding technique as described in section 2.2, then starting from (66) yields in parallel to (6),

$$P_{e|m} \leq \left(\sum_{m' \neq m} \sum_{\underline{y}} p_N(\underline{y}|\underline{x}^m)^{\frac{1}{\rho}} \cdot \psi_N^m(\underline{y})^{1-\frac{1}{\rho}} \cdot \left(\frac{Q_N(\underline{y}|\underline{x}^{m'})}{Q_N(\underline{y}|\underline{x}^m)} \right)^{\lambda} \right)^{\rho} , \quad \lambda \geq 0,\ 0 \leq \rho \leq 1 \,, \tag{67}$$

where as in section 2.2, $\psi_N^m(\underline{y})$ is the normalized Duman and Salehi tilting measure. In parallel to (8), the DS2 bound with the un-normalized tilting measure reads:

$$P_{e|m} \leq \left(\sum_{\underline{y}} G_N^m(\underline{y}) \cdot p_N(\underline{y}|\underline{x}^m) \right)^{1-\rho}$$
$$\cdot \left(\sum_{m' \neq m} \sum_{\underline{y}} p_N(\underline{y}|\underline{x}^m) \cdot G_N^m(\underline{y})^{1-\frac{1}{\rho}} \cdot \left(\frac{Q_N(\underline{y}|\underline{x}^{m'})}{Q_N(\underline{y}|\underline{x}^m)} \right)^{\lambda} \right)^{\rho} , \quad 0 \leq \rho \leq 1 \,, \quad \lambda \geq 0 \,. \tag{68}$$

The advantage of this bound, as compared to the original Gallager mismatched bound (66), is as in the matched case: The possible utilization for specific codes and ensembles. This is demonstrated in Section 4.2 for a fully interleaved fading channel with a faulty channel state information.

5.2. Ensembles of random codes

In this subsection, we examine the bound for a random coding strategy. As in the matched regime, for a general distribution of codewords $q_N(\underline{x})$, the DS2 bound with the optimized tilting measure reconstructs the 1965 Gallager random coding bound. We therefore continue with the mismatched

Gallager bound (66) and restrict our attention to ensembles of codes where each codeword satisfies the inequality:

$$N\varepsilon - \delta \le \Gamma_N(\underline{x}^j) < N\varepsilon\,, \quad \delta > 0\,, \quad j = 1, 2, \ldots, M\,, \tag{69}$$

where $\Gamma_N(\underline{x}^j)$ stands for an arbitrary cost function involved in the transmission of the codeword $\underline{x}^j$ and ε is a positive constant, both to be specified. For a codebook satisfying the inequality (69) (for every codeword), we further loosen the bound (66) letting,

$$P_{e|m} \le \sum_{\underline{y}} p_N(\underline{y}|\underline{x}^m) \cdot \left[\sum_{m' \ne m} \left(\frac{Q_N(\underline{y}|\underline{x}^{m'})}{Q_N(\underline{y}|\underline{x}^m)} \right)^{\lambda} \cdot \left(\frac{e^{\Gamma_N(\underline{x}^{m'})}}{e^{\Gamma_N(\underline{x}^m)}} \right) \cdot e^{\delta} \right]^{\rho}\,, \quad \lambda, \rho \ge 0\,. \tag{70}$$

This is since for every value of j ($j = 1, 2, \ldots, M$), the inequality: $\frac{e^{\Gamma_N(\underline{x}^{m'})}}{e^{\Gamma_N(\underline{x}^m)}} \cdot e^{\delta} > 1$ holds for any codebook satisfying (69). Now, we select randomly and independently each codeword in such a codebook by the probability law $\alpha_N(\underline{x})$. The upper bound on the decoding error probability over this ensemble equals:

$$P_e \le \sum_{\underline{y}} \sum_{\underline{x}^m} \alpha_N(\underline{x}^m) p_N(\underline{y}|\underline{x}^m) \sum_{\underline{x}^1} \sum_{\underline{x}^2} \cdots \sum_{\underline{x}^{m-1}} \sum_{\underline{x}^{m+1}} \cdots \sum_{\underline{x}^M} \left\{ \prod_{\substack{i=1 \\ i \ne m}}^{M} \alpha_N(\underline{x}^i) \left(\sum_{j \ne m} \left(\frac{Q_N(\underline{y}|\underline{x}^j)}{Q_N(\underline{y}|\underline{x}^m)} \right)^{\lambda} \frac{e^{\Gamma_N(\underline{x}^j)}}{e^{\Gamma_N(\underline{x}^m)}} e^{\delta} \right)^{\rho} \right\}\,, \tag{71}$$

where $\lambda, \rho \ge 0$.

As indicated in (69), the measure $\alpha_N(\underline{x})$ satisfies:

$$\alpha_N(\underline{x}) \equiv 0 \quad \forall\, \underline{x} \text{ where } \Gamma_N(\underline{x}) \notin [N\varepsilon - \delta,\, N\varepsilon)\,. \tag{72}$$

In principle, the parameter δ introduced in (69) may depend on N. Invoking the Jensen inequality in (71) yields the following upper bound on the decoding error probability:

$$P_e \le (M-1)^{\rho} \cdot e^{\delta\rho} \cdot \sum_{\underline{y}} \sum_{\underline{x}} \alpha_N(\underline{x}) \cdot p_N(\underline{y}|\underline{x}) \cdot \left\{ \sum_{\underline{x}'} \alpha_N(\underline{x}') \cdot \left(\frac{Q_N(\underline{y}|\underline{x}')}{Q_N(\underline{y}|\underline{x})} \right)^{\lambda} \cdot \frac{e^{\Gamma_N(\underline{x}')}}{e^{\Gamma_N(\underline{x})}} \right\}^{\rho}\,, \tag{73}$$

where $0 \le \rho \le 1$, $\lambda \ge 0$, and the superscript of $\underline{x}^m$ was removed (as the bound is clearly invariant to the transmitted codeword). Let $q_N(\underline{x})$ be an arbitrary probability measure and we set,

$$\alpha_N(\underline{x}) = \frac{q_N(\underline{x})}{\mu_{\alpha}}\,, \tag{74}$$

where based on (69):

$$\mu_{\alpha} = \mu_{\alpha}(\Gamma, \delta, \varepsilon) = \sum_{\underline{x}\,:\,\Gamma_N(\underline{x}) \in [N\varepsilon - \delta, N\varepsilon)} q_N(\underline{x})\,. \tag{75}$$

The substitution of (74) into (73) yields:

$$P_e \leq \frac{(M-1)^\rho \cdot e^{\delta\rho}}{(\mu_\alpha)^{1+\rho}} \cdot \sum_{\underline{y}} \sum_{\underline{x}} q_N(\underline{x}) \cdot p_N(\underline{y}|\underline{x}) \cdot \left\{ \sum_{\underline{x}'} q_N(\underline{x}') \cdot \left(\frac{Q_N(\underline{y}|\underline{x}')}{Q_N(\underline{y}|\underline{x})} \right)^\lambda \cdot \frac{e^{\Gamma_N(\underline{x}')}}{e^{\Gamma_N(\underline{x})}} \right\}^\rho , \qquad (76)$$

where $\lambda \geq 0$, $0 \leq \rho \leq 1$ and $\delta > 0$. We may further loosen the bound (76) by replacing $\frac{e^{\delta\rho}}{(\mu_\alpha)^{1+\rho}}$ by $\frac{e^\delta}{\mu_\alpha^2}$, as $\delta > 0$, $0 \leq \rho \leq 1$ and $\mu_\alpha \leq 1$. This yields:

$$P_e \leq \frac{e^\delta}{\mu_\alpha^2} M^\rho \cdot \sum_{\underline{y}} \sum_{\underline{x}} q_N(\underline{x}) \cdot p_N(\underline{y}|\underline{x}) \cdot \left\{ \sum_{\underline{x}'} q_N(\underline{x}') \cdot \left(\frac{Q_N(\underline{y}|\underline{x}')}{Q_N(\underline{y}|\underline{x})} \right)^\lambda \cdot \frac{e^{\Gamma_N(\underline{x}')}}{e^{\Gamma_N(\underline{x})}} \right\}^\rho , \qquad (77)$$

where we have also upper bounded $M-1$ by M. Let the channel probability law $p_N(\underline{y}|\underline{x})$ and the mismatched metric $Q_N(\underline{y}|\underline{x})$ be memoryless:

$$p_N(\underline{y}|\underline{x}) = \prod_{k=1}^N p(y_k|x_k),\ Q_N(\underline{y}|\underline{x}) = \prod_{k=1}^N Q(y_k|x_k) , \qquad (78)$$

and we also set an i.i.d. probability measure $q_N(\underline{x})$:

$$q_N(\underline{x}) = \prod_{k=1}^N q(x_k) . \qquad (79)$$

Let the cost function $\Gamma_N(\underline{x})$ be an additive function:

$$\Gamma_N(\underline{x}) = \sum_{k=1}^N \gamma(x_k) . \qquad (80)$$

Substituting (78)-(80) into (77) yields the single letter expression,

$$P_e \leq \frac{e^\delta}{\mu_\alpha^2} \cdot e^{\rho RN} \cdot \left[\sum_y \sum_x q(x) \cdot p(y|x) \cdot \left(\sum_{x'} q(x') \cdot \left(\frac{Q(y|x')}{Q(y|x)} \right)^\lambda \cdot \frac{e^{\gamma(x')}}{e^{\gamma(x)}} \right)^\rho \right]^N , \qquad (81)$$

where the rate R equals $R = \frac{\ln(M)}{N}$. Alternatively,

$$P_e \leq \frac{e^\delta}{\mu_\alpha^2} \cdot e^{-N[E_0(q,\gamma,\rho,\lambda)-\rho R]} , \qquad (82)$$

where

$$E_0(q,\gamma,\rho,\lambda) = -\ln \left[\sum_y \sum_x q(x) \cdot p(y|x) \cdot \left(\sum_{x'} q(x') \cdot \left(\frac{Q(y|x')}{Q(y|x)} \right)^\lambda \cdot \frac{e^{\gamma(x')}}{e^{\gamma(x)}} \right)^\rho \right] . \qquad (83)$$

Before turning to treat the exponent in (82), let us estimate μ_α for fixed (N-independent) δ. Based on (75) and (80):

$$\mu_\alpha = \text{Prob}\left(-\delta \leq \sum_{k=1}^{N} \gamma(x_k) - N\varepsilon < 0\right) , \tag{84}$$

where $x_k,\ k = 1, 2, \ldots, N$, are i.i.d. random variables governed by the single-letter probability measure $q(x)$. Now, we choose ε to satisfy:

$$\varepsilon = \sum_x q(x) \cdot \gamma(x) , \tag{85}$$

for a fixed $\gamma(x)$. Similarly to Gallager [16], one can estimate μ_α for $N \to \infty$, yielding

$$\lim_{N\to\infty} \sqrt{N} \cdot \mu_\alpha = \frac{\delta}{\sqrt{2\pi\sigma^2}} , \tag{86}$$

where

$$\sigma^2 = \sum_x q(x) \cdot \gamma(x)^2 - \left(\sum_x q(x) \cdot \gamma(x)\right)^2 . \tag{87}$$

This is directly found based on (84) by the central limit property as follows:

$$\begin{aligned} \mu_\alpha \quad &= \quad \text{Prob}\left(\frac{-\delta}{\sqrt{N}} \leq \frac{1}{\sqrt{N}} \cdot \sum_{k=1}^{N} \gamma(x_k) - \sqrt{N}\varepsilon < 0\right) \\ \underset{N\to\infty}{\longrightarrow} \quad & \quad \text{Prob}\left(\frac{-\delta}{\sqrt{N}} \leq N(0, \sigma^2) < 0\right) \underset{\frac{\delta}{\sqrt{N}} \ll 1}{\approx} \frac{\delta}{\sqrt{N}} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} . \end{aligned} \tag{88}$$

The pre-exponent factor in (82) can be optimized over $\delta \geq 0$ to yield,

$$\min_{\delta>0} \lim_{N\to\infty} \frac{1}{N} \frac{e^\delta}{\mu_\alpha^2} = \min_{\delta>0} (2\pi\sigma^2) \cdot \frac{e^\delta}{\delta^2} = (2\pi\sigma^2) \cdot \frac{e^2}{4} , \tag{89}$$

which demonstrates that the behavior of the pre-exponent in (82) is asymptotically proportional to N (with $N \to \infty$). Thus for finite ε (85) and σ^2 (87), the pre-exponent in (82) has no exponential implications as it behaves asymptotically like a $\frac{\ln(N)}{N}$ term in the exponent. For a fixed input distribution $q(x)$, we then attempt to maximize the error exponent

$$E(R, q) = \sup_{\substack{0\leq\rho\leq 1 \\ \lambda\geq 0 \\ \gamma(x)}} \left(E_0(q, \gamma, \rho, \lambda) - \rho R\right) , \tag{90}$$

where $\gamma(x)$ is a real function yielding finite ε and σ^2 in (85) and (87) respectively. Consider the rate equation:

$$R = \frac{\partial}{\partial \rho}\Big(E_0(q,\gamma,\rho,\lambda)\Big) = e^{E_0(q,\gamma,\rho,\lambda)} \tag{91}$$

$$\cdot \left\{ -\sum_y \sum_x q(x)\cdot p(y|x)\cdot \left(\sum_{x'} q(x')\cdot \left(\frac{Q(y|x')}{Q(y|x)}\right)^{\lambda} \cdot \frac{e^{\gamma(x')}}{e^{\gamma(x)}}\right)^{\rho} \cdot \ln\left[\frac{\sum_{x'} q(x')\cdot Q^{\lambda}(y|x')\cdot e^{\gamma(x')}}{Q^{\lambda}(y|x)\cdot e^{\gamma(x)}}\right]\right\},$$

where the last equality results from (83).
The maximal rate R_H is calculated by substituting $\rho = 0$ in (91), yielding:

$$R_H = \sup_{\gamma(x)} \max_{\lambda \geq 0} \sum_y \sum_x q(x)\cdot p(y|x)\cdot \ln\left(\frac{Q^{\lambda}(y|x)\cdot e^{\gamma(x)}}{\sum_{x'} q(x')\cdot Q^{\lambda}(y|x')\cdot e^{\gamma(x')}}\right), \tag{92}$$

This in fact equals what is known as Csiszár-Körner-Hui lower bound on the mismatched capacity, since it yields the dual representation in the terminology of [18]. It is rather straightforwardly verified that the error exponent $E(R)$ in (90) is positive for $R < R_H$. The input distribution $q(x)$ can now be chosen to maximize R_H, in the usual sense [18], [23]. In the case of continuous input and output alphabets, the relevant sums should be replaced by integrals and it should be verified that the optimized $\gamma(x)$ and $q(x)$ yield finite (though arbitrary) ε and σ in (85) and (87) respectively.

It can be also verified that at $\rho = 1$

$$R = \sup_{\lambda \geq 0} E_0(q, \gamma = 0\,,\, \rho = 1, \lambda)\,, \tag{93}$$

yields the generalized cutoff rate [19], which cannot be further improved by optimizing over the real function $\gamma(x)$. In terms of ensembles, it is worth emphasizing that the upper bound (82) resulted by restricting our attention to ensembles of codewords satisfying (69). This restriction is necessary, as if a purely random ensemble is attempted, then it was concluded in [18] that the results associated with $\gamma(x) = 0$ cannot be surpassed. The maximal rate then corresponds to the generalized mutual information (GMI) [23].

In [25], the Gallager 1965 bounding technique was applied to derive upper bounds for i.i.d. random and fixed composition codes, operating over memoryless channels with finite input alphabets and arbitrary output alphabets, and whose decoding metric is matched to the channel. The metric which is at the beginning general, was then optimized to yield an equivalent metric to ML decoding, and yet the results in [25] reproduced the error exponent for fixed composition codes [6]. Here we deal with a mismatched metric and the code ensemble is restricted by introducing a generalized energy constraint (69), which is subject to optimization.

5.3. Ensembles of structured codes

We apply here the generalization of the DS2 bound for the mismatched decoding regime (66) to ensembles of structured codes. We assume that the transition probabilities of the channel $p_N(\underline{y}|\underline{x})$ and the mismatched metric $Q_N(\underline{y}|\underline{x})$ are binary-input, symmetric-output and memoryless.

The optimization of the normalized Duman and Salehi tilting measure $\psi_N^m(\underline{y})$ in (67) is restricted here to the case where $\psi_N^0(\underline{y})$ can be expressed in the product form: $\psi_N^0(\underline{y}) = \prod_{i=1}^{N} \psi(y_i)$. In that case, the partitioning of the code (or ensemble of codes) to constant Hamming weight subcodes yields (37), where:

$$P_{e|0}(d) \leq (S_d)^\rho \cdot \left(\sum_y g_1(y) \cdot \psi(y)^{1-\frac{1}{\rho}} \right)^{(N-d)\rho} \cdot \left(\sum_y g_2(y) \cdot \psi(y)^{1-\frac{1}{\rho}} \right)^{d\rho} , \tag{94}$$

and

$$g_1(y) = p(y|0)^{\frac{1}{\rho}} , \quad g_2(y) = p(y|0)^{\frac{1}{\rho}} \cdot \left(\frac{Q(y|1)}{Q(y|0)} \right)^\lambda . \tag{95}$$

For continuous output channels, the sums in (94) should be replaced by integrals. With the aid of the calculus of variations, the optimal normalized Duman and Salehi tilting measure ψ, in terms of minimizing the upper bound (94)-(95), admits the form:

$$\psi(y) = \frac{p(y|0) \cdot \left(1 + \alpha \cdot \left(\frac{Q(y|1)}{Q(y|0)} \right)^\lambda \right)^\rho}{\sum_y p(y|0) \cdot \left(1 + \alpha \cdot \left(\frac{Q(y|1)}{Q(y|0)} \right)^\lambda \right)^\rho} , \quad \text{where } \lambda \geq 0 , \ 0 \leq \rho \leq 1 . \tag{96}$$

As demonstrated in Appendix B, the parameter α is optimally determined by a numerical solution of the equation:

$$\frac{\sum_y p(y|0) \cdot \left(1 + \alpha \cdot \left(\frac{Q(y|1)}{Q(y|0)} \right)^\lambda \right)^{\rho-1}}{\sum_y p(y|0) \cdot \left(1 + \alpha \cdot \left(\frac{Q(y|1)}{Q(y|0)} \right)^\lambda \right)^\rho} = 1 - \delta , \tag{97}$$

where $\delta \equiv \frac{d}{N}$ is the normalized Hamming weight of the N-length codewords possessing a Hamming weight d. The existence and uniqueness of a solution α in (97) is proved in Appendix B.

From the discussion above, the upper bound for the mismatched decoding regime in (37), (94)-(97) involves numerical optimizations over the two parameters $\lambda \geq 0$, $0 \leq \rho \leq 1$, for at most the $N+1$

constant Hamming weight subcodes of the considered ensemble of codes. The necessary information on the ensemble of codes comprises the average distance spectrum (or even a tight upper bound on the ensemble distance spectrum), and a refined information on the algebraic structure of the codes is not required.

6. Some Applications and Examples

In Section 5, a large class of efficient recent bounds (or their Chernoff versions) was demonstrated to be a special case of the generalized DS2 bound. Implications and applications of these observations are pointed out here, including the fully interleaved fading channel, resorting to either matched or mismatched decoding. The proposed approach can be also generalized to geometrically uniform non-binary codes, finite state channels, bit interleaved coded modulation systems, and it can be also used for the derivation of upper bounds on the conditional decoding error probability.

6.1. AWGN channels

We apply here some variants of the Gallager bounds and other reported bounds to the ensemble of un-punctured turbo codes of rate $\frac{1}{3}$ with a uniform interleaver of length 1000 and two recursive systematic convolutional (RSC) component codes whose generators are $(1, \frac{21}{37})$ in octal form (see Fig. 1). The considered ensemble of codes is also terminated to the all-zero state at the end of each frame by additional four bits (having thus an overall of 3012 coded bits). The following upper bounds on the bit error probability are depicted in Fig. 2: The original tangential sphere bound of Poltyrev ([24],[30]) (differing from the loosened Chernoff version bound derived in [7]), Engdahl and Zigangirov bound [13] (the non-Chernoff version in Section 4.6 here), Duman and Salehi bounds [11, 12] (Sections 2.3 and 4.2 here), Divsalar bound [7] (Section 4.5 here), Viterbi and Viterbi bounds [38, 39] (Sections 4.3 and 4.4 here), and finally the ubiquitous union bound. It is demonstrated in Fig. 2, that the difference between the two versions of the Duman and Salehi bounds for the binary-input AWGN channel is very small (about 0.01 dB) and that also the Duman and Salehi (first version) bound coincides with Divsalar bound. The first observation is consistent with [12] and the second observation verifies numerically the fact that Divsalar bound is indeed a closed form of Duman and Salehi (first version) bound [11], as was first indicated in [7] (the negligible difference between the two bounds depicted in Fig. 2 is attributed to the numerical optimizations associated with the calculation of the first version of the Duman and Salehi bound (41)). The tangential sphere bound [24] and Engdahl and Zigangirov bound [13] were derived without invoking the Chernoff bounding technique, explaining therefore their advantage over Duman and Salehi bounds and some other related bounds for block codes of moderate block-length, as the latter bounds rely on the Chernoff bounding technique (see Fig. 2). However, as indicated in [7] (and section 4.7), the loosening of the tangential sphere bound which turns it into a particular case of the DS2 bound, does not carry any implications on its associated error exponent for the asymptotically infinite block length.

6.2. Fully interleaved fading channels

We apply here various variants of the generalized DS2 bound to the fully interleaved Rayleigh fading channel with perfect channel state information (see Section 4.3). These bounds are evaluated for the ensemble of rate $\frac{1}{3}$ uniformly interleaved turbo codes depicted in Fig. 1. The optimization of the generalized DS2 bound (combined with the tight version of the union bound) is demonstrated as the tightest reported bound (see also [34],[26]). It also approximately replicates here the performance of these turbo codes when iteratively decoded with 10 iterations of the Log-MAP iterative decoding algorithm (see Fig. 3). These bounds are also evaluated for the ensemble of rate-$\frac{1}{4}$ uniformly interleaved RA codes [9], with a block length of $N = 4096$ (see Figs. 4,5). In [34], [26] these bounds were applied to some ensembles of efficient codes (turbo [5], Gallager-LDPC [17] and RA codes [9]). For moderate values of energy per bit to spectral noise density $\left(\frac{E_b}{N_0}\right)$, the optimized DS2 upper bound (under ML decoding) falls below the computer simulation results for the sum-product iterative decoding algorithm (see Fig. 5), demonstrating the mild sub-optimality of the latter decoding algorithm.

6.3. Thresholds of codes

Let C be an ensemble of binary, linear (or geometrically uniform) block codes of block-length N and rate R. The threshold is defined as the minimal value of $\frac{E_b}{N_0}$ ensuring a block decoding error probability that decays to zero for asymptotically infinite block length. Obviously, the threshold depends on three parameters: The channel model, the applied decoding algorithm and the considered code (or ensemble of codes). The thresholds associated with some decoding algorithms were recently studied: an analytical upper bound on the thresholds of ML decoded block codes operating over the binary-input AWGN channel was derived by Divsalar [7]. Upper bounds on the thresholds of RA [9] and Gallager-LDPC codes [17] under "typical pairs" decoding algorithm, were derived by McEliece et. al. [1] for the binary symmetric channel (BSC) and for the binary-input AWGN channel. Exact thresholds for the BSC under a specific iterative decoding algorithm were derived by Bazzi et. al. [3]. Based on the optimized DS2 bound, an upper bound on the thresholds which are associated with ML decoding were numerically calculated for the fully interleaved fading channels with perfect channel state information (see [26], section 3.3). As predicted, these thresholds meet the ultimate Shannon capacity limit for the ensemble of fully random block codes of rate R. Thresholds for the ensembles of RA and (j,k) Gallager-LDPC codes are depicted in Table 1. These thresholds refer to the fully interleaved Rayleigh fading channel with perfect channel state information and a maximum ratio combining (MRC) space diversity of order four. As expected, due to the simple structure of RA codes, the calculated thresholds for the ensemble of RA codes are worse than the corresponding thresholds for the ensembles of Gallager-LDPC codes. In general, the calculation of the upper bounds on these thresholds depends on the asymptotic exponent of the distance spectrum of the considered codes (or ensembles). Recently, some techniques for the calculation of the asymptotic exponents of the distance spectra of turbo codes and irregular LDPC codes where derived in [43] and [44], respectively. These new techniques can be applied to the calculation of upper bounds on the $\frac{E_b}{N_o}$-thresholds for turbo and LDPC codes under optimal ML decoding.

6.4. Mismatched decoding

We apply here the generalized DS2 bound introduced in section 5.3, to study the robustness of a mismatched decoder that is based on ML decoding with respect to the faulty channel measurements. We examine here a BPSK modulated signal, transmitted through a fully interleaved Rayleigh fading channel. For simplicity, we apply our bounds to the case of a perfect phase estimation of the i.i.d. fading samples (in essence reducing the problem to a real channel). We also assume that the estimated and real magnitudes of the Rayleigh fading samples satisfy a joint distribution of two correlated bivariate Rayleigh variables with an average power of unity. We note that if the phase estimation is not perfect (as opposed to what we have assumed here, that the error occurs only in the amplitude estimation), assuming that the actual fading coefficient (a) and its estimate $(\hat{a})$ are correlated Gaussian circularly symmetric random variables, the implication of the correlation decoder manifest itself in just some enhanced additive independent Gaussian noise. This is in fact the best (matched) decoder given the perfect side information at hand. This is since:

$$y = ax + n = \hat{a}x + (a - \hat{a})x + n.$$

In the Gaussian regime the estimated error $(a - \hat{a})$ is Gaussian distributed and independent of the estimator $\hat{a}$, even when conditioned on x (as neither the mean nor the variance of $(a - \hat{a})x$ does not depend on x for the binary case, as $||x||^2 = E_s$ [42]). The perfectly known phase treated here is a different case, since the results are not captured in terms of scaling the additive noise.

Based on the notations in section 5, the binary-input and symmetric-output channel is memoryless, and we therefore obtain:

$$\begin{aligned} Q(y, \hat{a}|0) &= Q(-y, \hat{a}|1) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left[-\frac{1}{2} \cdot \left(y + \hat{a}\sqrt{\frac{2RE_b}{N_0}}\right)^2\right] \cdot 2\hat{a} \cdot \exp(-\hat{a}^2)\,, \\ p(y, \hat{a}|0) &= p(-y, \hat{a}|1) = \int_0^{+\infty} \frac{1}{\sqrt{2\pi}} \cdot \exp\left[-\frac{1}{2} \cdot \left(y + a\sqrt{\frac{2RE_b}{N_0}}\right)^2\right] \cdot p_{a,\hat{a}}(a, \hat{a})\, da\,, \\ & \qquad -\infty < y < +\infty\,,\ \hat{a} \geq 0\,, \end{aligned} \tag{98}$$

where $a, \hat{a}$ are jointly Rayleigh distributed:

$$p_{a,\hat{a}}(r_1, r_2) = \frac{4r_1r_2}{1-\rho^*} \cdot I_0\left(\frac{2\sqrt{\rho^*}\, r_1r_2}{1-\rho^*}\right) \cdot \exp\left(-\frac{r_1^2 + r_2^2}{1-\rho^*}\right)\,, \quad r_1,\, r_2 \geq 0\,, \tag{99}$$

$E(a^2) = E(\hat{a}^2) = 1$, and the parameter ρ^* designates the correlation coefficient between the pairs of squared Rayleigh random variables $a^2, \hat{a}^2$:

$$\rho^* = \frac{cov\,(a^2,\, \hat{a}^2)}{\sqrt{Var(a^2) \cdot Var(\hat{a}^2)}}\,. \tag{100}$$

The integral expressed in (98),(99) can be transformed to another integral which is easily calculated with the aid of the Gaussian numerical integration formula (see Appendix C).

By partitioning the considered linear and binary block code C of length N and rate R to constant Hamming weight subcodes, the tight version of the union bound on the ML decoding error probability, corresponding to the subcode of Hamming weight d ($0 \leq d \leq N$), is given in this case by the expression:

$$S_d \cdot E\left[Q\left(\sqrt{\frac{2RE_b}{N_0}}\, \frac{\sum_{i=1}^{d} a_i\, \hat{a}_i}{\sqrt{\sum_{i=1}^{d} \hat{a}_i^2}}\right)\right]\,, \tag{101}$$

where S_d designates the number of codewords of Hamming weight d, the Q-function is introduced in (53) and the notation E stands for the statistical expectation with respect to the i.i.d. Rayleigh fading samples $\{\underline{a}_i\}$ and their Rayleigh distributed estimations $\{\underline{\hat{a}}_i\}$. The expressions in (101) (where $0 \leq d \leq N$) are calculated via the Monte-Carlo method, by generating the correlated bivariate Rayleigh random variables with a certain correlation coefficient ρ^* (expressed in (100)), based on the algorithm proposed in [37]. The upper bound on the bit error probability based on the generalized DS2 bound in (94)-(97) is compared in Fig. 6 to the improved upper bound that combines the DS2 bound with the tight version of the union bound for every constant Hamming weight subcode. The comparison between these two bounds refers to the ensemble of uniformly interleaved turbo codes depicted in Fig. 1 where $\frac{E_b}{N_0} = 3$ dB. It is reflected from Fig. 6 that the latter bound yields a considerably tighter error floor (the error floor in Figs. 6,7 is observed for a sufficiently high correlation coefficient ρ^*, and it reflects the robustness of the system to faulty measurements of the fading samples). The error floor exhibited by the improved upper bound in Fig. 6 also predicts reliably the error floor of the turbo codes associated with the suboptimal and efficient Log-MAP iterative decoding algorithm (based on computer simulations with 10 iterations). The upper bounds on the bit error probability that are based on the combination of the generalized DS2 bound in (94)-(97) and the tight version of the union bound (101) are depicted in Fig. 7 for several fixed values of $\frac{E_b}{N_0}$. These curves are plotted as a function of the correlation coefficient ρ^* between the squares of the i.i.d. jointly Rayleigh fading samples and their estimates. The bounds in Fig. 7 refer to the ensemble of unpunctured rate $\frac{1}{3}$ turbo codes depicted in Fig. 1. Since for a fully interleaved Rayleigh fading channel with *perfect* side information on the fading samples, the channel cutoff rate corresponds to $\frac{E_b}{N_0} = 3.23$ dB, then according to the upper bounds depicted in Fig. 7, the ensemble performance of these turbo codes (associated with the hypothetical ML decoding) is sufficiently robust in case of mismatched decoding, even in a portion of the rate region exceeding the channel cutoff rate.

The proposed upper bounds depicted in Figs. 6 and 7 were efficiently implemented in Matlab software, indicating their applicability in terms of complexity and the practical running time involved in their calculations.

7. Summary and Conclusions

In this paper, we discuss numerous efficient bounds on the decoding error probability of specific codes and ensembles under maximum-likelihood (ML) decoding, and we demonstrate the underlying connections that exist between them. In addressing the Gallager bounding techniques and their variations, we focus here on the Duman and Salehi variation, which originates from the classical 1965 Gallager bound. The considered upper bounds on the block and bit error probabilities under ML decoding rely on the distance spectrum and the input-output weight distributions of the codes, respectively, which are in general calculable. By generalizing the second version of the Duman and Salehi (DS2) bound which was originally derived for the binary-input AWGN channel [12], we demonstrate its remarkable generality and show that it provides the natural bridge between the 1961 and 1965 Gallager bounds (see Section 3). It is applicable for both random and specific codes, as well as for either bit or block error analysis. Some observations and interconnections between the Gallager and Duman and Salehi bounds are presented in Section 3, which partially rely on insightful observations made by Divsalar [7]. In particular, it is demonstrated in Section 3 that the

1965 Gallager random coding bound can be re-derived from the DS2 bound. The geometric interpretations of these Gallager-type bounds are introduced in Section 3, reflecting the non-uniqueness of their associated probability tilting measures. In Section 4, many reported upper bounds on the ML decoding error probability (or their Chernoff versions) are shown to be special cases of the DS2 bound. This is done by choosing an appropriate probability tilting measure and calculating the resulting bound with the specific chosen measure. This framework also facilitates to generalize the Shulman and Feder bound [36] (see Appendix A).

The proposed approach can be generalized to geometrically uniform non-binary codes, finite state channels, bit interleaved coded modulation systems, and it can be also used for the derivation of upper bounds on the conditional decoding error probability (as to account for a possible partitioning of the original code to subcodes). The DS2 bound is generalized in Section 5 to the mismatched decoding regime, where the decoder performs ML decoding with respect to a mismatched metric. The generalized bound is applicable to the analysis of the error performance of deterministic codes and ensembles. We address in particular the random coding version, which matches the Gallager 1965 random coding setting, and reproduces in Section 5 some known results (see [18], [19], [23]), hinging on an alternative approach which appropriately limits the considered ensemble of codes. Implications and applications of these observations are pointed out in Section 6, which include the fully interleaved fading channel with either matched or mismatched decoding.

The generalization of the DS2 upper bound for the analysis of the bit error probability yields the replacement of the distance spectrum $\{S_d\}_{d=1}^{N}$ (which appears in the upper bound on the block error probability) by the sequence $\{S'_d\}_{d=1}^{N}$ where: $S'_d = \sum_{w=1}^{NR} \left\{ \left(\frac{w}{NR}\right) A_{w,d} \right\}$, and $A_{w,d}$ designates the number of codewords in a systematic block code with an information Hamming weight w and a total Hamming weight d (where $0 \leq w \leq NR$ and $0 \leq d \leq N$). The derivation of the generalized DS2 bound on the bit error probability is detailed in [26], and it yields a considerable improvement in the tightness of the bound, as compared to [11, 12]. The computational complexity which is involved in calculating the DS2 bound is moderate.

The interconnections between many reported upper bounds are depicted in Fig. 8, where it is reflected that the DS2 bound yields many reported upper bounds as special cases.

The tangential sphere bound [24] happens often to be the tightest reported upper bound for block codes which are transmitted over the binary-input AWGN channel and ML decoded (see Fig. 2 and [30, 31, 32, 41]). However, in the random coding setting, it fails to reproduce the random coding error exponent (see [24]), while the DS2 bound does. In fact, also Shulman and Feder bound which is a special case of the latter bound (see section 4.1) achieves capacity for the ensemble of fully random block codes. This substantiates the claim that there is no uniformly best bound, providing the incentive for the generalization of Shulman and Feder bound (see Appendix A). However, we note that the loosened version of the tangential sphere bound [7] (which involves the Chernoff inequality) maintains the asymptotic (i.e, for infinite block length) exponential tightness of the tangential sphere bound of Poltyrev [24], and this loosened version is a special case of the DS2 bound (see Section 4.2.6).

Acknowledgements:
The authors wish to thank Robert Gallager, Emre Telatar, Rüdiger Urbanke and the anonymous reviewers for their many constructive suggestions for sharpening the meaning, and enhancing the lucidity of the presentation of this paper.

Appendix A: A generalization of Shulman and Feder Bound

In this appendix a generalized version of Shulman and Feder bound is derived, based on the DS2 bound.

Let C be the ensemble of random binary block codes of length N and rate $R = \frac{\ln(M)}{N}$ nats per channel use, where the codewords are independently selected and uniformly distributed:

$$q_N(\underline{x}) = \prod_{i=1}^{N} q(x_i) = 2^{-N} \, . \tag{A.1}$$

We assume here that an arbitrary codeword $\underline{x}$ is transmitted over a binary-input output-symmetric channel. The upper bound (11) on the conditional decoding error probability is independent of the transmitted codeword (under the assumption (A.1)), which can then be on one hand averaged with respect to the pure binomial distribution (giving the 1965 Gallager random coding bound) and it is also valid if $\underline{x}$ is chosen arbitrarily, say the all zero codeword. For this specific choice, the Gallager bound on the conditional decoding error probability yields

$$P_{e|0} \leq (M-1)^{\rho} \cdot \sum_{\underline{y}} p_N(\underline{y}|\underline{0}) \cdot \left(\sum_{\underline{x}'} 2^{-N} \left(\frac{p_N(\underline{y}|\underline{x}')}{p_N(\underline{y}|\underline{0})} \right)^{\lambda} \right)^{\rho} , \quad \lambda \geq 0 \, , \; 0 \leq \rho \leq 1 \, . \tag{A.2}$$

The inner summation in (A.2) can be alternatively rewritten as a double summation, where its outer summation is over the Hamming weights of the N-length binary vectors and its inner summation is over the binary vectors of the same Hamming weight, giving:

$$\sum_{\underline{x}'} 2^{-N} \cdot \left(\frac{p_N(\underline{y}|\underline{x}')}{p_N(\underline{y}|\underline{0})} \right)^{\lambda} = \sum_{l=0}^{N} 2^{-N} \sum_{j=1}^{\binom{N}{l}} \left(\frac{p_N(\underline{y}|\underline{\varepsilon}^{l,j})}{p_N(\underline{y}|\underline{0})} \right)^{\lambda} \tag{A.3}$$

where $\varepsilon^{l,j}$ stands for the j-th permutation of an N-length binary vector of a Hamming weight l (where $l = 0, 1, \ldots, N, \quad j = 1, 2, \ldots, \binom{N}{l}$). The substitution of (A.3) into (A.2) gives:

$$P_{e|0} \leq (M-1)^{\rho} \cdot \sum_{\underline{y}} p_N(\underline{y}|\underline{0}) \cdot \left(\sum_{l=0}^{N} 2^{-N} \cdot \binom{N}{l} \cdot \frac{1}{\binom{N}{l}} \cdot \sum_{j=1}^{\binom{N}{l}} \left(\frac{p_N(\underline{y}|\underline{\varepsilon}^{l,j})}{p_N(\underline{y}|\underline{0})} \right)^{\lambda} \right)^{\rho} . \tag{A.4}$$

Let $b_l = 2^{-N} \cdot \binom{N}{l}$ $(l = 0, 1, \ldots, N)$ and we designate by E_l the statistical expectation over the $\binom{N}{l}$ binary N-length vectors of Hamming weight l. Equation (A.4) then transforms to

$$P_{e|0} \leq (M-1)^{\rho} \cdot \sum_{\underline{y}} p_N(\underline{y}|\underline{0})^{1-\lambda\rho} \cdot \left(\sum_{l=0}^{N} b_l \cdot E_l \left(p_N(\underline{y}|\underline{\varepsilon}^{l})^{\lambda} \right) \right)^{\rho} , \; \lambda \geq 0 \, , \; 0 \leq \rho \leq 1 \, . \tag{A.5}$$

Upon substitution of $\lambda = \frac{1}{1+\rho}$, invoking the assumption of a memoryless binary-input and symmetric-output channel, the right hand side of equation (A.5) yields the 1965 Gallager random coding bound, i.e:

$$(M-1)^{\rho} \cdot e^{-N \cdot E\left(R, q=\frac{1}{2}\right)} , \tag{A.6}$$

where $E_0(\rho, q)$ is introduced in (3).

The formulation (A.5) of the 1965 Gallager random coding bound allows for the derivation of a generalized version of the Shulman and Feder bound [36] as a particular case of the generalized DS2 bound. To that end, let C be a fixed binary linear block code, where its distance spectrum is designated by $\{S_l\}_{l=0}^{N}$. We consider here the ensemble of codes including the codebook of the original block code and also all its possible position permutations. Since all the codes in this ensemble feature the same ML decoding error probability, then the conditional decoding error probability (where without any loss of generality, the all-zero vector is assumed to be the transmitted codeword) does not depend on the codebook selected from this ensemble. Based on (8), we obtain the upper bound:

$$\begin{aligned} P_{e|0} \quad &\leq \quad \left(\sum_{\underline{y}} G_N^0(\underline{y}) \cdot p_N(\underline{y}|\underline{0}) \right)^{1-\rho} \\ &\cdot \quad E\left(\left(\sum_{m' \neq 0} \sum_{\underline{y}} p_N(\underline{y}|\underline{0}) \cdot G_N^0(\underline{y})^{1-\frac{1}{\rho}} \cdot \left(\frac{p_N(\underline{y}|\underline{x}^{m'})}{p_N(\underline{y}|\underline{0})} \right)^{\lambda} \right)^{\rho} \right) , \end{aligned} \tag{A.7}$$

$$0 \leq \rho \leq 1 \, , \ \lambda \geq 0 \, ,$$

where E designates the statistical expectation over all the codebooks of the considered ensemble of codes, and $G_N^0(\underline{y})$ is assumed to be independent of the selected codebook in the ensemble (as it will be later verified).

By invoking the Jensen inequality, we further loosen the upper bound (A.7) and obtain:

$$\begin{aligned} P_{e|0} \quad &\leq \quad \left(\sum_{\underline{y}} G_N^0(\underline{y}) \cdot p_N(\underline{y}|\underline{0}) \right)^{1-\rho} \cdot \\ &\cdot \quad \left(\sum_{\underline{y}} p_N(\underline{y}|\underline{0}) \cdot G_N^0(\underline{y})^{1-\frac{1}{\rho}} \cdot \sum_{m' \neq 0} E\left(\left(\frac{p_N(\underline{y}|\underline{x}^{m'})}{p_N(\underline{y}|\underline{0})} \right)^{\lambda} \right) \right)^{\rho} , \quad \begin{array}{c} 0 \leq \rho \leq 1 \\ \lambda \geq 0 \, . \end{array} \end{aligned} \tag{A.8}$$

Since the statistical expectation in (A.8) is taken over all the codebooks of the considered ensemble, it therefore only depends on the Hamming weight of the binary N-length vector $\underline{x}^{m'}$. Thus, the upper bound (A.8) can be expressed in terms of the ensemble distance spectrum:

$$P_{e|0} \leq \left(\sum_{\underline{y}} G_N^0(\underline{y}) \cdot p_N(\underline{y}|\underline{0}) \right)^{1-\rho}$$

$$\cdot \left(\sum_{\underline{y}} p_N(\underline{y}|\underline{0}) \cdot G_N^0(\underline{y})^{1-\frac{1}{\rho}} \cdot \sum_{l=0}^{N} S_l \cdot E_l \left(\left(\frac{p_N(\underline{y}|\underline{\varepsilon}^l)}{p_N(\underline{y}|\underline{0})} \right)^{\lambda} \right) \right)^{\rho} , \qquad (A.9)$$

$$0 \leq \rho \leq 1 \, , \ \lambda \geq 0 \, ,$$

where the Hamming weight of the N-length binary vector $\underline{\varepsilon}^l$ is $l(l = 0, 1, \ldots, N)$. By selecting the (non-negative) function $G_N^0(\underline{y})$ in (A.9) as:

$$G_N^0(\underline{y}) = \left(\sum_{l=0}^{N} S_l \cdot E_l \left(\left(\frac{p_N(\underline{y}|\underline{\varepsilon}^l)}{p_N(\underline{y}|\underline{0})} \right)^{\lambda} \right) \right)^{\rho} , \qquad (A.10)$$

the following upper bound results:

$$P_{e|0} \leq \sum_{\underline{y}} p_N(\underline{y}|\underline{0}) \cdot \left(\sum_{l=0}^{N} S_l \cdot E_l \left(\left(\frac{p_N(\underline{y}|\underline{\varepsilon}^l)}{p_N(\underline{y}|\underline{0})} \right)^{\lambda} \right) \right)^{\rho} , \qquad (A.11)$$

$$\lambda \geq 0 \, , \ 0 \leq \rho \leq 1 \, .$$

Let $v_l = \frac{S_l}{M-1}$ $(l = 0, 1 \ldots N)$, where $M-1$ is the number of non-zero codewords differing from the transmitted codeword, be a probability measure, and set b_l to be the binomial measure in (A.4). Hölder's inequality for an arbitrary non-negative sequence $\{Q_l\}_{l=0}^{N}$ yields:

$$\sum_{l=0}^{N} v_l \cdot Q_l = \sum_{l=0}^{N} b_l \cdot \left(\frac{v_l}{b_l} \right) \cdot Q_l \qquad (A.12)$$

$$\leq \left(\sum_{l=0}^{N} b_l \cdot Q_l^P \right)^{\frac{1}{P}} \cdot \left(\sum_{l=0}^{N} b_l \cdot \left(\frac{v_l}{b_l} \right)^{Q} \right)^{\frac{1}{Q}} , \quad \frac{1}{P} + \frac{1}{Q} = 1 \, , \ P, Q \geq 1 \, .$$

When invoked in (A.11), Hölder's inequality gives:

$$\begin{aligned}
P_{e|0} &\le (M-1)^{\rho} \cdot \sum_{\underline{y}} p_N(\underline{y}|\underline{0}) \cdot \left(\sum_{l=0}^{N} v_l \cdot E_l \left(\left(\frac{p_N(\underline{y}|\underline{\varepsilon}^l)}{p_N(\underline{y}|\underline{0})} \right)^{\lambda} \right) \right)^{\rho} \\
&\le (M-1)^{\rho} \cdot \left(\sum_{l=0}^{N} b_l \cdot \left(\frac{v_l}{b_l} \right)^{Q} \right)^{\frac{\rho}{Q}} \cdot \sum_{\underline{y}} p_N(\underline{y}|\underline{0}) \cdot \left(\sum_{l=0}^{N} b_l \cdot \left(E_l \left(\left(\frac{p_N(\underline{y}|\underline{\varepsilon}^l)}{p_N(\underline{y}|\underline{0})} \right)^{\lambda} \right) \right)^{P} \right)^{\frac{\rho}{P}} \\
&= (M-1)^{\rho} \left(\sum_{l=0}^{N} b_l \left(\frac{v_l}{b_l} \right)^{Q} \right)^{\frac{\rho}{Q}} \cdot \sum_{\underline{y}} p_N(\underline{y}|\underline{0})^{1-\lambda\rho} \cdot \left(\sum_{l=0}^{N} b_l \cdot \left(E_l \left(\left(p_N(\underline{y}|\underline{\varepsilon}^l) \right)^{\lambda} \right) \right)^{P} \right)^{\frac{\rho}{P}} \quad \text{(A.13)}
\end{aligned}$$

Invoking again the Jensen inequality $\left(E(\beta)\right)^{P} \le E(\beta^P)$ for $P \ge 1$, establishes the upper bound:

$$P_{e|0} \le (M-1)^{\rho} \cdot \sum_{\underline{y}} p_N(\underline{y}|\underline{0})^{1-\lambda\rho} \cdot \left(\sum_{l=0}^{N} b_l \cdot E_l \left(p_N(\underline{y}|\underline{\varepsilon}^l) \right)^{\lambda P} \right)^{\frac{\rho}{P}} \cdot \left(\sum_{l=0}^{N} b_l \cdot \left(\frac{v_l}{b_l} \right)^{Q} \right)^{\frac{\rho}{Q}} . \quad \text{(A.14)}$$

Suppose that the binary-input symmetric-output channel is memoryless:

$$\begin{aligned}
p_N(\underline{y}|\underline{x}) &= \prod_{i=1}^{N} p(y_i|x_i) , \\
p(y|0) &= p(-y|1) .
\end{aligned} \quad \text{(A.15)}$$

By introducing in (A.14) the new parameters λ', ρ' which satisfy the equalities:

$$\lambda' = \lambda P , \ \rho' = \frac{\rho}{P} , \text{ and } \lambda' = \frac{1}{1+\rho'} \quad \left(\lambda' \ge 0 , \ 0 \le \rho' \le \frac{1}{P} \right) , \quad \text{(A.16)}$$

the bound in (A.14) in view of (A.6) yields:

$$\begin{aligned}
P_{e|0} \le (M-1)^{\rho' P} &\cdot e^{-N \cdot E_0\left(\rho', q=\frac{1}{2}\right)} \cdot \left(\sum_{l=0}^{N} b_l \cdot \left(\frac{v_l}{b_l} \right)^{Q} \right)^{\frac{P\rho'}{Q}} \\
&\le e^{NRP\rho'} \cdot e^{-N \cdot E_0\left(\rho', q=\frac{1}{2}\right)} \cdot e^{\frac{\rho' P}{Q} \cdot \ln\left(\sum_{l=0}^{N} b_l \cdot \left(\frac{v_l}{b_l} \right)^{Q} \right)} \\
&= e^{-N \cdot E\left(\rho', PR + \frac{1}{N} \cdot \frac{P}{Q} \cdot \ln\left(\sum_{l=0}^{N} b_l \cdot \left(\frac{v_l}{b_l} \right)^{Q} \right), q=\frac{1}{2}\right)} , \quad 0 \le \rho' \le \frac{1}{P} , \quad \text{(A.17)}
\end{aligned}$$

where $E(\rho, R, q)$ is introduced in Eq. (4).

The upper bound (A.17) is a generalization of the Shulman and Feder bound [36], where the latter follows by letting $P = 1$, $Q \to \infty$, and noticing that:

$$\lim_{Q\to\infty} \frac{1}{N} \cdot \frac{1}{Q} \cdot \ln\left(\sum_{l=0}^{N} b_l \cdot \left(\frac{v_l}{b_l}\right)^Q\right) = \frac{1}{N} \cdot \max_{0\le l\le N} \left(\frac{v_l}{b_l}\right) . \tag{A.18}$$

The generalized Shulman and Feder bound is a particular case of the DS2 bound for the class of memoryless binary-input and output-symmetric channels. The generalized bound (A.17) is expected to improve the tightness of Shulman and Feder bound for codes whose distance spectrum deviates from the binomial distribution especially at low Hamming weights (e.g., turbo and LDPC codes). For a fully random block code whose distance spectrum is the binomial distribution, $P = 1$ is optimal in (A.17) and the Shulman and Feder bound matches the 1965 Gallager random coding bound. Considering the term in the error exponent of (A.17):

$$\begin{aligned}
&\frac{1}{N} \cdot \frac{P}{Q} \cdot \ln\left(\sum_{l=0}^{N} b_l \cdot \left(\frac{v_l}{b_l}\right)^Q\right) &&\overset{\text{a}}{=} \\
&\frac{P-1}{N} \cdot \ln\left(\sum_{l=0}^{N} v_l \cdot \left(\frac{v_l}{b_l}\right)^{\frac{1}{P-1}}\right) &&\overset{\text{b}}{\geq} \\
&\frac{P-1}{N} \cdot \sum_{l=0}^{N} v_l \cdot \ln\left(\left(\frac{v_l}{b_l}\right)^{\frac{1}{P-1}}\right) &&= \\
&\frac{1}{N} \cdot \sum_{l=0}^{N} v_l \cdot \ln\left(\frac{v_l}{b_l}\right) &&= \\
&\frac{1}{N} \cdot D(\underline{v}||\underline{b}) \ ,
\end{aligned} \tag{A.19}$$

where $D(\underline{v}||\underline{b})$ stands for the divergence between the two probability measures $\underline{v}$ and $\underline{b}$. Equality (a) in (A.19) follows by the relations: $\frac{P}{Q} = P - 1$, $Q = 1 + \frac{1}{P-1}$, while inequality (b) in (A.19) is due to the Jensen inequality: $E\Big(\ln(\beta)\Big) \le \ln\Big(E(\beta)\Big)$). The divergence in the right hand side of inequality (A.19) constitutes a lower bound on the rate loss with respect to pure random codes (with a binomial distance spectrum), and hence it provides some quantitative support to Battail's proposition [2] for the design of weakly random-like turbo codes.

Appendix B: On the Derivation of the Bound for the Mismatched Decoding Metric

Based on the variation of calculus technique, we derive the normalized Duman and Salehi tilting measure that minimize the upper bound (94)-(95). Repeating the steps in [26] for the mismatched decoding regime and setting $\alpha = \frac{k_2}{k_1}$ (where the constants k_1, k_2 are introduced in Appendix A in [26]), we then obtain the optimal tilting measure (96) in terms of minimizing the bound (94)-(95). Similarly to the calculations in Appendix A in [26], the parameter α in (96) satisfies the implicit equation:

$$\alpha = \frac{\delta}{1-\delta} \cdot \frac{\sum_y p(y|0) \cdot \left(1 + \alpha \cdot \left(\frac{Q(y|1)}{Q(y|0)}\right)^\lambda\right)^{\rho-1}}{\sum_y p(y|0) \cdot \left(\frac{Q(y|1)}{Q(y|0)}\right)^\lambda \cdot \left(1 + \alpha \cdot \left(\frac{Q(y|1)}{Q(y|0)}\right)^\lambda\right)^{\rho-1}} \,. \tag{B.1}$$

Equation (B.1) can be alternatively expressed as:

$$\begin{aligned} \frac{1-\delta}{\delta} &= \frac{\sum_y p(y|0) \cdot \left(1 + \alpha \cdot \left(\frac{Q(y|1)}{Q(y|0)}\right)^\lambda\right)^{\rho-1}}{\sum_y p(y|0) \cdot \left(1 + \alpha \cdot \left(\frac{Q(y|1)}{Q(y|0)}\right)^\lambda\right)^{\rho} - \sum_y p(y|0) \cdot \left(1 + \alpha \cdot \left(\frac{Q(y|1)}{Q(y|0)}\right)^\lambda\right)^{\rho-1}} \\ &= \frac{1}{\frac{\sum_y p(y|0) \cdot \left(1 + \alpha \cdot \left(\frac{Q(y|1)}{Q(y|0)}\right)^\lambda\right)^{\rho}}{\sum_y p(y|0) \cdot \left(1 + \alpha \cdot \left(\frac{Q(y|1)}{Q(y|0)}\right)^\lambda\right)^{\rho-1}} - 1} \,. \end{aligned} \tag{B.2}$$

Equation (B.2) immediately yields (97). We now prove the existence and uniqueness of a non-negative solution α for (97): To that end, we define the function:

$$t \equiv 1 + \alpha \cdot \left(\frac{Q(y|1)}{Q(y|0)}\right)^\lambda \,, \quad (\lambda \geq 0) \tag{B.3}$$

and the optimization equation (97) then transforms to the form:

$$\frac{E_y(t^{\rho-1})}{E_y(t^\rho)} = 1 - \delta \,, \quad (0 \leq \rho \leq 1) \,, \tag{B.4}$$

where E_y denotes the statistical expectation with respect to y, which is governed by the probability density function $p(y|0)$. Based on the alternative representation (B.4) of (97), we prove the existence and uniqueness of a solution to the implicit equation (B.1).

Existence: For $\alpha = 0$, t in (B.3) equals 1 and therefore also the left hand side of (B.4) equals 1. If $\alpha \gg 1$, then based on (B.3): $t \approx \alpha \cdot \left(\frac{Q(y|1)}{Q(y|0)}\right)^\lambda$, and therefore the asymptotic behavior of the left hand side in (B.4) is proportional to $\frac{1}{\alpha}$ (that obviously tends to zero for $\alpha \to +\infty$. As the left hand side in (B.4) is a continuous function of the parameter α, then it admits every value between 0 and 1 for non-negative values of α. Since the normalized Hamming distance δ lies in the interval $[0, 1]$, then the right hand side of (B.4) lies in the same interval. Therefore, for every such value of δ, there exists a non-negative value of α satisfying (B.4).

Uniqueness: Since the function t in (B.3) is increasing in the parameter α (for non-negative values of α) and also $0 \leq \rho \leq 1$, then the numerator and denominator of the left hand side in (B.4) are decreasing and increasing functions of α respectively (where $\alpha \geq 0$). Being also positive for non-negative values of α, the quotient in the left hand side of (B.4) is a decreasing function of α. The uniqueness of the solution α in equation (B.4) follows immediately. □

Appendix C: A Simplification of the Integral in (98),(99)

The integral in (98),(99) can be rewritten in the form:

$$p(y, \hat{a}|0) = \int_0^\infty \frac{1}{\sqrt{2\pi}} \cdot \exp\left[-\frac{(y+ac)^2}{2}\right] \cdot 4a\,\hat{a}\,\alpha\, I_0(Da)\, \exp\Big[-\alpha(a^2+\hat{a}^2)\Big]\, da\ , \qquad \text{(C.1)}$$

where:

$$c = \sqrt{\frac{2RE_b}{N_0}}\ ,\ \alpha = \frac{1}{1-\rho}\ ,\ D = \frac{2\sqrt{\rho}\,\hat{a}}{1-\rho}\ .$$

By a straightforward algebra, we get:

$$p(y, \hat{a}|0) = 4\alpha\,\hat{a}\, \exp\left[-\left(\alpha\,\hat{a}^2 + \frac{y^2}{2}\right)\right] \cdot \int_0^\infty \frac{a}{\sqrt{2\pi}} \cdot \exp\left[-\frac{1}{2}\,(\beta^2 a^2 + 2acy)\right] \cdot I_0(Da)\, da\ , \qquad \text{(C.2)}$$

where $\beta^2 = C^2 + 2\alpha$.

Substituting in (C.2) the integral representation of the modified Bessel function of the zero order:

$$I_0(x) = \frac{1}{2\pi} \int_0^{2\pi} e^{-x\cos\theta}\, d\theta\ , \qquad \text{(C.3)}$$

and interchanging the order of integration yields:

$$p(y, \hat{a}|0) = 4\alpha\,\hat{a}\, \exp\left[-\left(\alpha\,\hat{a}^2 + \frac{y^2}{2}\right)\right] \cdot \frac{1}{2\pi} \int_0^{2\pi} \int_0^\infty \frac{a}{\sqrt{2\pi}} \cdot \exp\left[-\frac{1}{2}\,(\beta^2 a^2 + 2acy)\right] \cdot \exp(-Da\cos\theta)\, da\, d\theta\ . \qquad \text{(C.4)}$$

The internal integral in (C.4) can be expressed in a closed form:

$$\int_0^{\infty} \frac{a}{\sqrt{2\pi}} \cdot \exp\left[-\frac{1}{2}\left(\beta^2 a^2 + 2acy\right)\right] \cdot \exp(-Da\cos\theta)\, da = \frac{1}{\sqrt{2\pi}\,\beta^2} - \frac{\gamma}{\beta^3} \exp\left(\frac{\gamma^2}{2\beta^2}\right) Q\left(\frac{\gamma}{\beta}\right) \;, \quad \text{(C.5)}$$

where $\gamma \triangleq \gamma(\theta) = cy + D\cos\theta$, and Q is introduced in (53).

By substituting (C.5) into (C.4), we obtain:

$$p(y, \hat{a}|0) = \left(\frac{2\alpha\,\hat{a}}{\beta^2}\right) \cdot \exp\left[-\left(\alpha\,\hat{a}^2 + \frac{y^2}{2}\right)\right] \cdot \left[\sqrt{\frac{2}{\pi}} - \frac{1}{\pi}\int_0^{2\pi} \delta(\theta)\, \exp\left(\frac{\delta^2(\theta)}{2}\right) \cdot Q\Big(\delta(\theta)\Big)\, d\theta\right] \;, \quad \text{(C.6)}$$

where $\delta(\theta) = \frac{\gamma(\theta)}{\beta}$.

Equality (C.6) can be alternatively rewritten as:

$$p(y, \hat{a}|0) = \left(\frac{2\sqrt{2}\alpha\,\hat{a}}{\beta^2}\right) \exp\left[-\left(\alpha\,\hat{a}^2 + \frac{y^2}{2}\right)\right] \cdot \left(\frac{1}{\sqrt{\pi}} - \frac{1}{\pi}\int_0^{\pi} \phi(\theta)\, erfcx\Big(\phi(\theta)\Big)\, d\theta\right) \;, \quad \text{(C.7)}$$

where:

$$erfcx\,(x) = e^{x^2}\, erfc\,(x) = 2\, e^{x^2}\, Q(\sqrt{2}\,x) \quad -\infty < x < +\infty \;, \quad \text{(C.8)}$$

and

$$\phi(\theta) = \frac{\delta(\theta)}{\sqrt{2}} \qquad 0 \leq \theta < 2\pi \;. \quad \text{(C.9)}$$

The integral in the right hand side of (C.7) is numerically calculable by the Gaussian numerical integration formula.

References

[1] S. Aji, H. Jin, A. Khandekar, R. J. McEliece and D. J. C. Mackay, "BSC thresholds for code ensembles based on 'typical pairs' decoding", in the book *Codes, Systems and Graphical Models*, Springer-Verlag Series IMA Volumes in Mathematics and its Applications (B. Marcus and J. Rosental, eds.), vol. 123, pp. 195–210, 2001.

[2] G. Battail, "A conceptual framework for understanding turbo codes", *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 2, pp. 245–254, February 1998.

[3] L. Bazzi, T. Richardson and R. Urbanke, "Exact thresholds and optimal codes for the binary symmetric channel and Gallager's decoding algorithm A", accepted to *IEEE Transactions on Information Theory.* [Online]. Available: *http://lthcwww.epfl.ch/publications.html.*

[4] E. R. Berlekamp, "The technology of error correction codes", *Proceedings of the IEEE*, vol. 68, no. 5, pp. 564–593, May 1980.

[5] C. Berrou, A. Glavieux and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding", *Proceedings 1993 IEEE International Conference on Communications (ICC '93)*, pp. 1064–1070, Geneva, Switzerland, May 1993.

[6] I. Csiszár and J. Körner, Information Theory: Coding Theorems for Discrete Memoryless Systems, Academic Press, NY 1981.

[7] D. Divsalar, "A simple tight bound on error probability of block codes with application to turbo codes", TMO Progress Report 42–139, JPL, pp. 1–35, November 1999.

[8] D. Divsalar and E. Biglieri, "Upper bounds to error probabilities of coded systems over AWGN and fading channels", *Proceedings 2000 IEEE Global Telecommunications Conference (GLOBECOM'00)*, pp. 1605–1610, San Francisco, California, USA, November 2000.

[9] D. Divsalar, H. Jin and R. J. McEliece, "Coding theorems for 'turbo-like' codes", *Proceedings of the 1998 Allerton Conference*, Monticello, Illinois, pp. 201–210, September 1998.

[10] S. Dolinar, L. Ekroot and F. Pollara, "Improved error probability bounds for block codes on the Gaussian channel", *Proceedings 1994 IEEE International Symposium on Information Theory*, Trondheim, Norway, June 24–July 1, 1994.

[11] T. M. Duman and M. Salehi, "New performance bounds for turbo codes", *IEEE Trans. on Communications*, vol. 46, no. 6, pp. 717–723, June 1998.

[12] T. M. Duman, "Turbo codes and turbo coded modulation systems: Analysis and performance bounds", Ph.D. dissertation, Elect. Comput. Eng. Dep., Northeastern University, Boston, MA USA, May 1998.

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

[14] R. M. Fano, Transmission of Information, jointly published by the M.I.T Press and John Wiley & Sons, 1961.

[15] R. G. Gallager, "A simple derivation of the coding theorem and some applications", *IEEE Trans. on Information Theory*, vol. 11, no. 1, pp. 3–18, January 1965.

[16] R. G. Gallager, Information Theory and Reliable Communications, John Wiley, 1968.

[17] R. G. Gallager, *Low-Density Parity-Check Codes*, Cambridge, MA USA, MIT Press, 1963.

[18] A. Ganti, A. Lapidoth and I. E. Telatar, "Mismatched decoding revisited: general alphabets, channels with memory and the wide-band limit", *IEEE Trans. on Information Theory*, vol. 46, no. 7, pp. 2315–2327, November 2000.

[19] G. Kaplan and S. Shamai, "Information rates and error exponents of compound channels with application to antipodal signaling in a fading environment", *AEU*, vol. 47, no. 4, pp. 228–239, 1993.

[20] A. Lapidoth and S. Shamai, "Fading channels: How perfect need "perfect side-information" be ? ", *IEEE Trans. on Information Theory*, vol. 48, no. 5, pp. 1118–1130, May 2002.

[21] D. J. C. Mackay and R. M. Neal, "Near Shannon limit performance of low-density parity-check codes", *IEEE Electronic Letters*, vol. 33, no. 6, pp. 457–458, March 1997.

[22] R. J. McEliece, "How to compute weight enumerators for convolutional codes", in *Communications and Coding*, M. Darnel and B. Honary, eds., New York: John Wiley & Sons, chapter 6, pp. 121–141, 1998.

[23] N. Merhav, G. Kaplan, A. Lapidoth and S. Shamai, "On information rates for mismatched decoders", *IEEE Trans. on Information Theory*, vol. 40, no. 6, pp. 1953–1967, November 1994.

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

[25] G. Poltyrev, "Random coding bounds for discrete memoryless channels", *Problems of Information Transmission*, vol. 18, no. 1, pp. 9–21, January-March, 1982; translated from Problemy Predachi Informatsii, vol. 18, no. 1, pp. 12-26 (Russian).

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

[27] I. Sason and S. Shamai, "Improved upper bounds on the ML performance of turbo codes for interleaved Rician fading channels, with comparison to iterative decoding", *Proceedings 2000 IEEE International Conference on Communications (ICC 2000)*, pp. 591–596, New Orleans, LA USA, 18–22 June 2000.

[28] I. Sason and S. Shamai, "On improved bounds on coded communications over interleaved fading channels, with application to turbo codes", *Proceedings of the Second International Symposium on Turbo Codes and Related Topics*, pp. 239–243, Brest, France, 4–7 September 2000.

[29] I. Sason and S. Shamai, "On Gallager bounding technique with applications to turbo-like codes over fading channels", *Proceedings 21st IEEE Convention of the Electrical and Electronic Engineers in Israel*, pp. 431–434, Tel-Aviv, Israel, 11–12 April 2000.

[30] I. Sason and S. Shamai, "Improved upper bounds on the ML decoding error probability of parallel and serial concatenated turbo codes via their ensemble distance spectrum", *IEEE Trans. on Information Theory*, vol. 46, no. 1, pp. 24–47, January 2000.

[31] I. Sason and S. Shamai, "Bounds on the error probability of ML decoding for block and turbo-block codes", *Annals of Telecommunications*, vol. 54, no. 3–4, pp. 183–200, March-April 1999.

[32] I. Sason and S. Shamai, "Improved upper bounds on the ensemble performance of ML decoded low-density parity-check codes", *IEEE Communications Letters*, vol. 4, no. 3, pp. 89–91, March 2000.

[33] I. Sason and S. Shamai, "Gallager's 1961 bound: extensions and observations", Technical Report, CC No. 258, Technion, Israel, October 1998.

[34] I. Sason, S. Shamai and D. Divsalar, "On simple and tight upper bounds on the ML decoding error probability for block codes over interleaved fading channels", *Proceedings Sixth International Symposium on Communication Theory and Applications (ISCTA'01)*, pp. 236–241, Ambleside, UK, 15–20 July 2001.

[35] S. Shamai and I. Sason, "Variations on Gallager's bounding technique with applications", *Proceedings of the 2nd International Symposium on Turbo Codes and Related Topics*, pp. 27–34, Brest, France, 4–7 September 2000.

[36] N. Shulman and M. Feder, "Random coding techniques for nonrandom codes", *IEEE Trans. on Information Theory*, vol. 45, no. 6, pp. 2101–2104, September 1999.

[37] C. Tellambura and A. D. S. Jayalath, "Generation of bivariate Rayleigh and Nakagami-m fading envelopes", *IEEE Communications Letters*, vol. 4, no. 5, pp. 170–172, May 2000.

[38] A. M. Viterbi and A. J. Viterbi, "Improved union bound on linear codes for the binary-input AWGN channel, with application to turbo codes", *Proceedings of the IEEE International Symposium on Information Theory (ISIT 1998)*, Cambridge, MA, p. 29, August 16–21, 1998.

[39] A. M. Viterbi and A. J. Viterbi, "An improved union bound for binary linear codes on the AWGN channel, with application to turbo decoding", *Proceedings of IEEE Information Theory Workshop*, p. 72, San Diego, California, February 1998.

[40] A. M. Viterbi and A. J. Viterbi, "New results on serial concatenated and accumulated convolutional turbo code performance", *Annals of Telecommunications*, vol. 54, no. 3–4, pp. 173–182, March-April 1999.

[41] I. Sason, "Upper bounds on the maximum likelihood decoding error probability for block codes and turbo-like codes", Ph.D. dissertation, Department of Electrical Engineering, Technion – Israel Institute of Technology, Haifa, Israel, September 2001. [Online]. Available: *http://lthiwww.epfl.ch/∼eeigal/*.

[42] H. Shin and J. H. Lee, "Improved upper bound on the bit error probability of turbo codes for ML decoding with perfect CSI in a Rayleigh fading channel", *Proceedings 2001 IEEE 12th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2001)*, San Diego, CA, USA, pp. A.169–A.173, September 2001.

[43] I. Sason, E. Telatar and R. Urbanke, "On the asymptotic input-output weight distributions and thresholds of convolutional and turbo-like codes", *IEEE Trans. on Information Theory*, vol. 48, no. 12, pp. 3052–3061, December 2002.

[44] D. Burshtein and G. Miller, "Asymptotic enumeration methods for analyzing LDPC codes", presented in the *40th Annual Conference on Communication, Control and Computing*, Allerton House, Monticello, Illinois, USA, October 2–4, 2002.

Figure Captions

Figure 1: Block diagram of an ensemble of uniformly interleaved turbo codes of rate-$\frac{1}{3}$ and interleaver length 1000. The generator of the two identical component codes (16–states, recursive systematic convolutional (RSC) codes) is $\left[1, \frac{1+D^4}{1+D+D^2+D^3+D^4}\right]$. A termination to the all-zero state is assumed at the end of each block.

Figure 2: A comparison between upper bounds on the bit error probability with ML decoding. The bounds refer to the ensemble of turbo codes in Fig. 1, operating over a binary-input AWGN channel.

1. Tangential sphere bound ([24], [30]).
2. Engdahl and Zigangirov bound [13].
3. DS2 bound [12].
4. Divsalar bound [7].
5. Duman and Salehi (first version) bound [11].
6. Viterbi & Viterbi (second version) bound [38].
7. Union bound in Q-form.

The bounds 2–6 are combined with the union bound in its Q-form for every constant Hamming-weight subcode of the considered ensemble of codes.

Figure 3: A comparison between upper bounds on the bit error probability with ML decoding. The bounds refer to the ensemble of turbo codes in Fig. 1, operating over a fully interleaved Rayleigh fading channel with perfect channel state information.

1. The generalized DS2 bound [26] combined with the union bound in its tight form.
2. The generalization of Engdahl and Zigangirov bound [26].
3. The union bound in its tight form.

These upper bounds are also compared to computer simulation results of the Log-MAP iterative decoding algorithm with up to 10 iterations.

Figure 4: Block diagram of the ensemble of rate-$\frac{1}{4}$ and uniformly interleaved RA codes. The 1024 information bits are repeated $q = 4$ times, then the 4096 symbols are uniformly interleaved and differentially encoded.

Figure 5: Comparison between upper bounds on the bit error probability with ML decoding. The bounds refer to the RA codes in Fig. 4, operating over a fully interleaved Rayleigh fading channel with perfect channel state information.

1. The generalized DS2 bound [26].

2. Divsalar and Biglieri bound [8].

3. The generalization of Engdahl and Zigangirov bound [26].

4. The generalization of Viterbi and Viterbi bound [34].

5. The tight form of the union bound.

The upper bounds 1–4 are combined with the tight version of the union bound for every constant Hamming-weight subcode of the considered ensemble of codes. The bounds are also compared with computer simulation results of the sum-product iterative decoding algorithm with 20 iterations.

Figure 6: A comparison between two upper bounds on the bit error probability. The bounds refer to the ensemble of turbo codes depicted in Fig. 1, operating over a fully interleaved Rayleigh fading channel with mismatched decoding. The generalized DS2 bound is compared with the improved bounding technique which combines the DS2 bound with the tight form of the union bound (for every constant Hamming weight subcode). The bounds are demonstrated for $\frac{E_b}{N_0} = 3$ dB, and are depicted as a function of the correlation coefficient (in the range $\frac{1}{2}$ to 1) between the squares of the i.i.d. jointly Rayleigh fading samples and their estimates.

Figure 7: A comparison between upper bounds on the bit error probability for the ensemble of turbo codes depicted in Fig. 1. The codes operate over a fully interleaved Rayleigh fading channel with mismatched decoding. The bounds are based on combining the generalized DS2 bound with the tight form of the union bound, and it is applied to every constant Hamming weight subcode. The bounds are plotted for $\frac{E_b}{N_0} = 2.50, 2.75, 3.00$ and 3.25 dB, as a function of the correlation coefficient (in the range $\frac{1}{2}$ to 1) between the squares of the i.i.d. jointly Rayleigh fading samples and their estimates.

Figure 8: An interconnections diagram among the various upper bounds.

Rate	$\frac{E_b}{N_0}$ **thresholds for the ensembles of** **RA codes**	**(j,k) Gallager-LDPC codes**	**The Shannon capacity limit**
$\frac{1}{2}$	-	(4,8) : 0.76 dB, (5,10) : 0.69 dB	0.58 dB
$\frac{1}{3}$	1.02 dB	(4,6) : -0.11 dB, (8,12) : -0.21 dB	-0.25 dB
$\frac{1}{4}$	0.12 dB	(3,4) : -0.29 dB, (6,8) : -0.58 dB	-0.62 dB
$\frac{1}{5}$	-0.30 dB	(4,5) : -0.73 dB	-0.82 dB

Table 1: Upper bounds on the $\frac{E_b}{N_o}$–thresholds for ensembles of codes operating over a fully interleaved Rayleigh fading channel with space diversity (based on the maximum ratio combining principle) of order four, and perfect channel state information at the receiver.

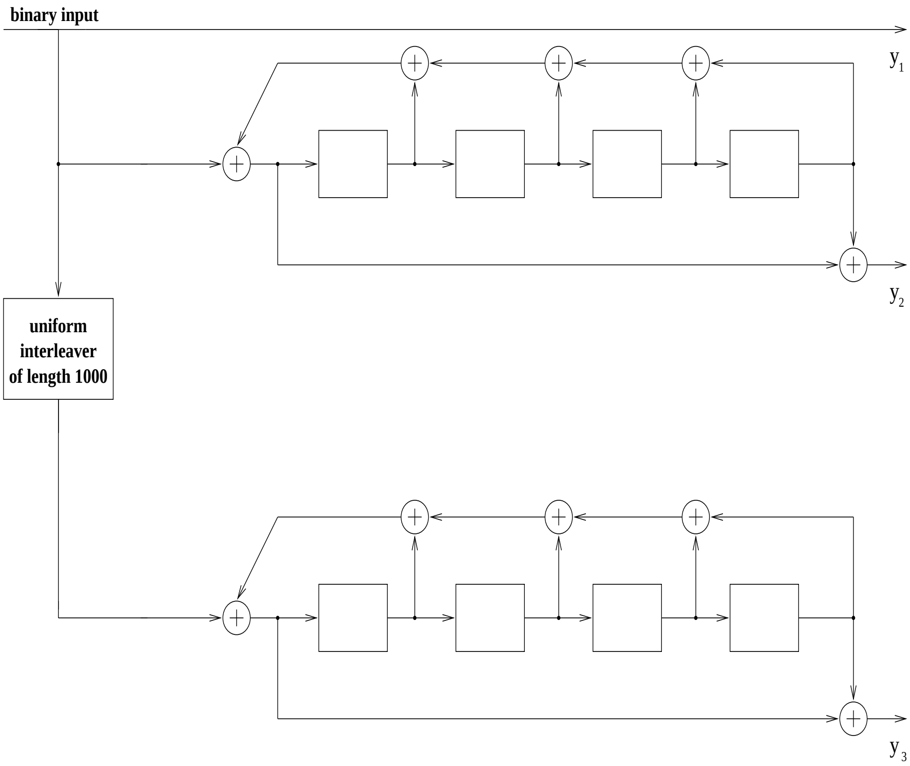


Figure 1.

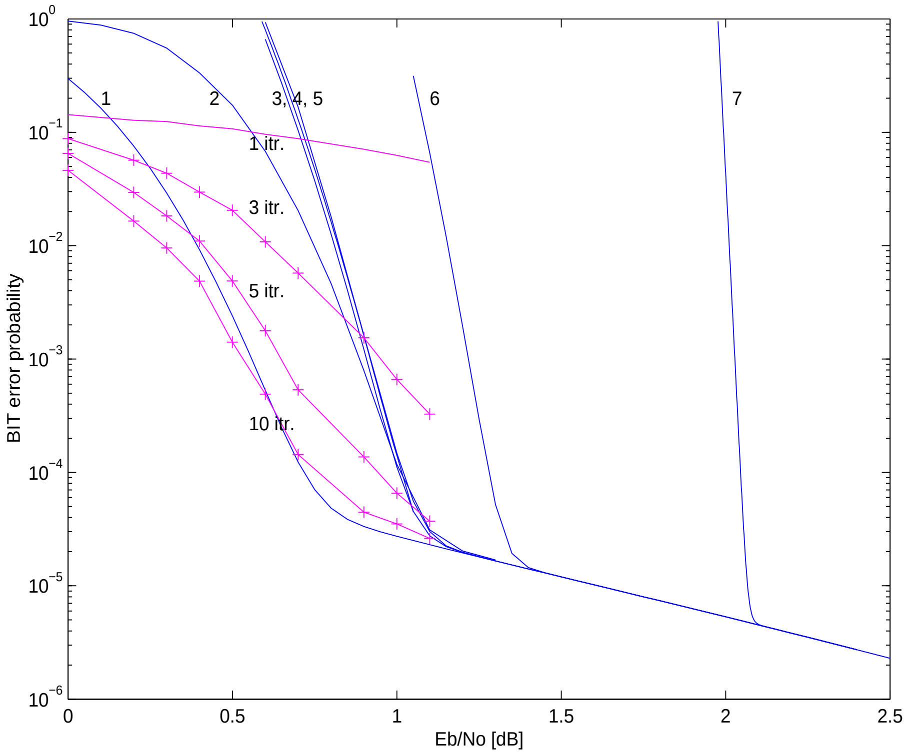
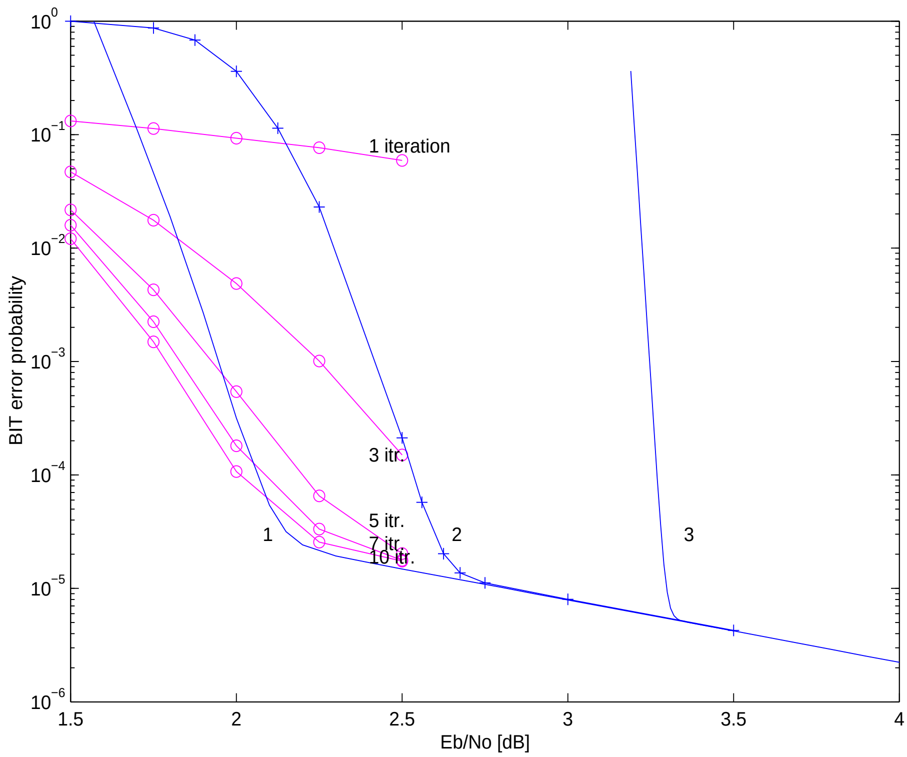


Figure 2.



BIT error probability
Eb/No [dB]
1 iteration
3 itr.
5 itr.
7 itr.
10 itr.
1
2
3

Figure 3.

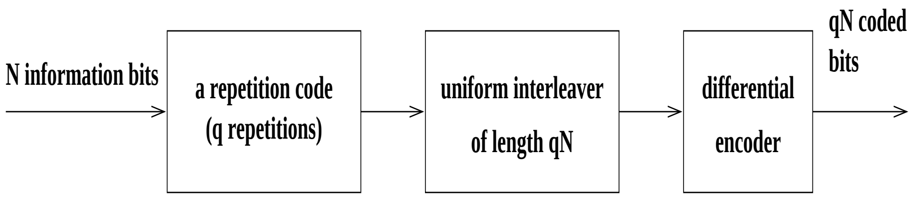


Figure 4.

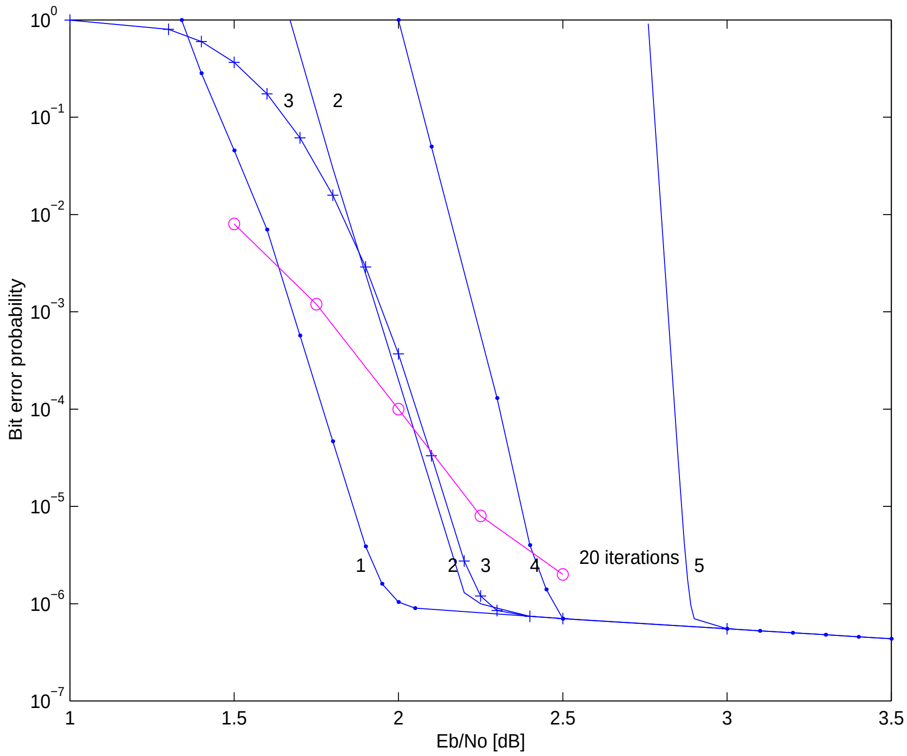


Figure 5.

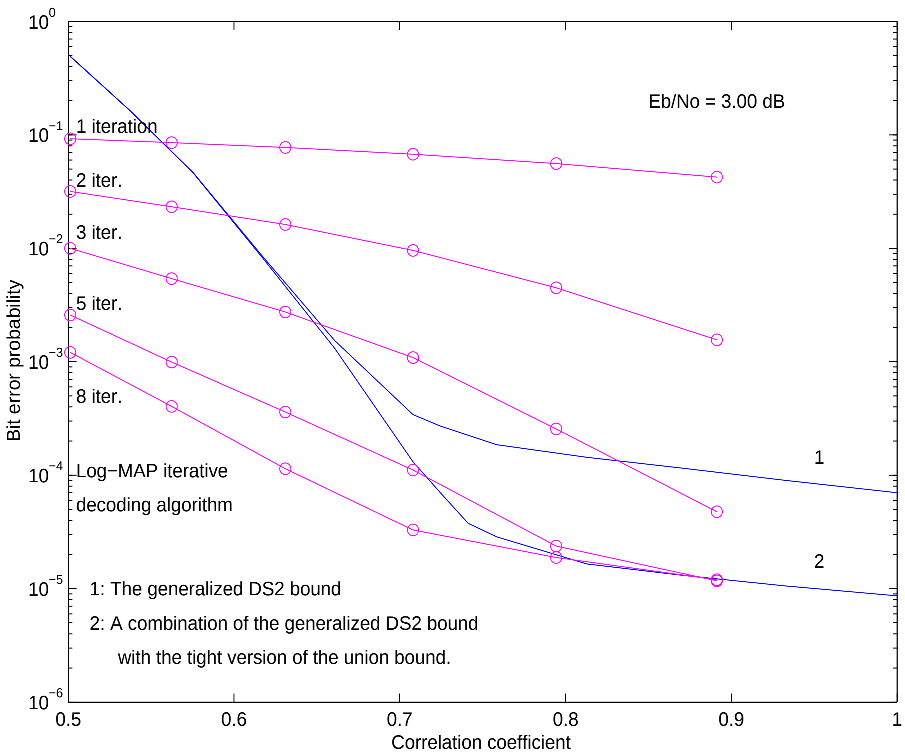


Figure 6.

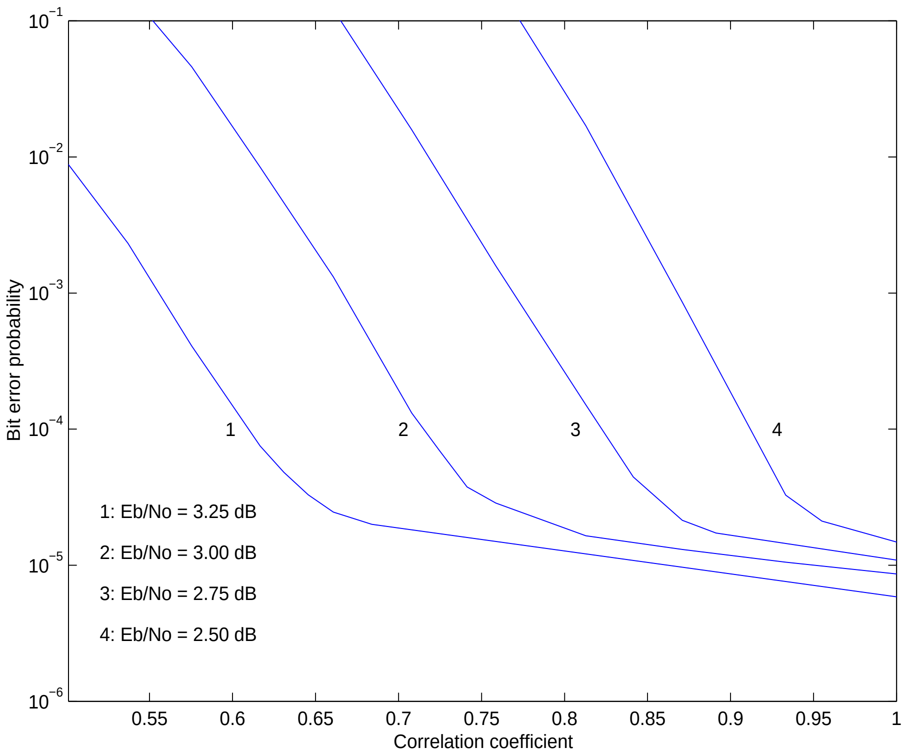
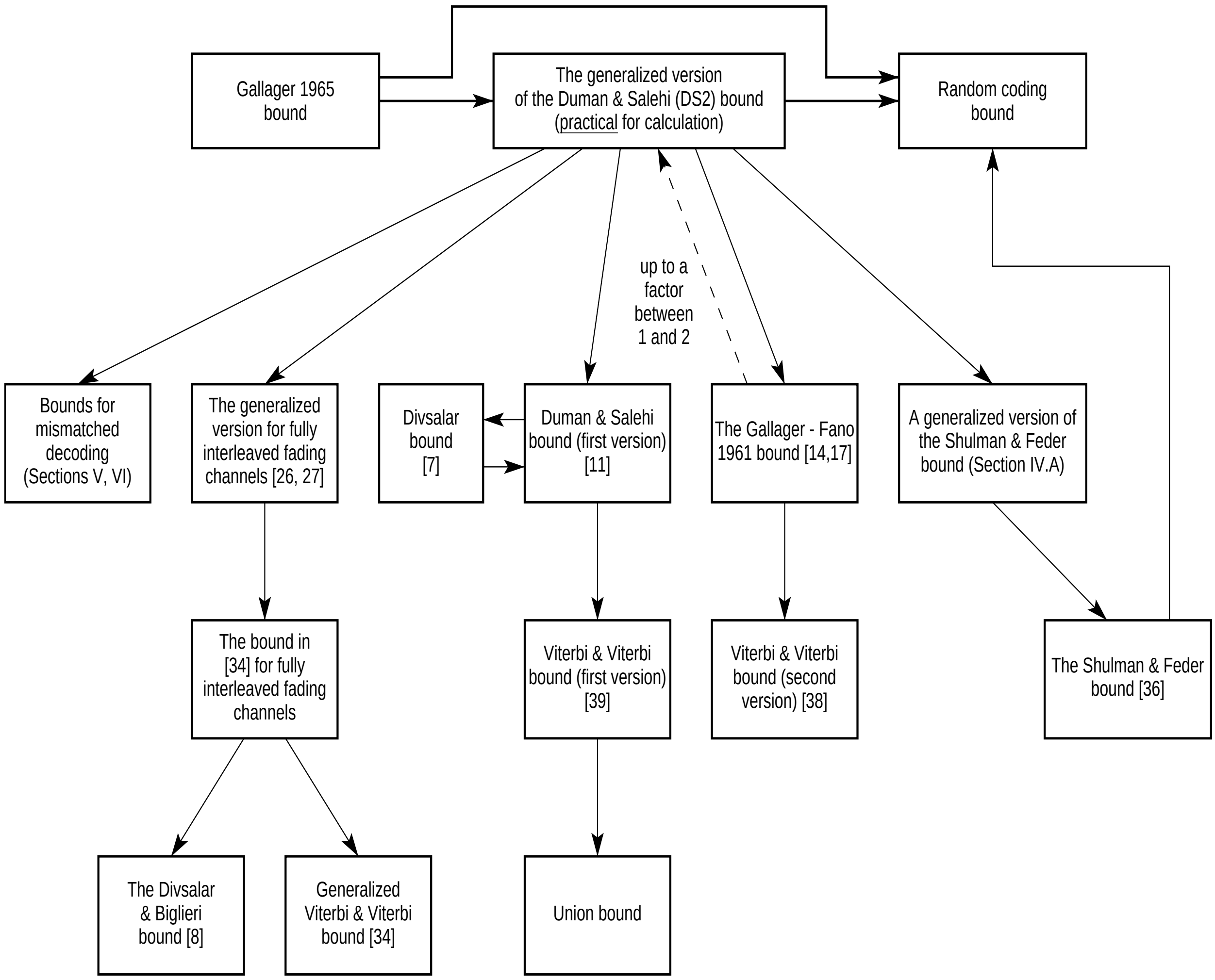


Figure 7.


Gallager 1965 bound
The generalized version of the Duman & Salehi (DS2) bound (practical for calculation)
Random coding bound
up to a factor between 1 and 2
Bounds for mismatched decoding (Sections V, VI)
The generalized version for fully interleaved fading channels [26, 27]
Divsalar bound [7]
Duman & Salehi bound (first version) [11]
The Gallager - Fano 1961 bound [14,17]
A generalized version of the Shulman & Feder bound (Section IV.A)
The bound in [34] for fully interleaved fading channels
Viterbi & Viterbi bound (first version) [39]
Viterbi & Viterbi bound (second version) [38]
The Shulman & Feder bound [36]
The Divsalar & Biglieri bound [8]
Generalized Viterbi & Viterbi bound [34]
Union bound