Coding for Parallel Channels: Gallager Bounds and Applications to Turbo-Like Codes

Igal Sason, *Member*, Idan Goldenberg, *Student Member*

Abstract

The transmission of coded communication systems is widely modeled to take place over a set of parallel channels. This model is used for transmission over block-fading channels, rate-compatible puncturing of turbo-like codes, multi-carrier signaling, multilevel coding etc. New upper bounds on the maximum-likelihood (ML) decoding error probability are derived in the parallel-channel setting. We focus on the generalization of the Gallager-type bounds and discuss the connections between some versions of these bounds. The tightness of these bounds for parallel channels is exemplified for structured ensembles of turbo codes, repeat-accumulate codes and some of their recent variations (e.g., punctured accumulate-repeat-accumulate codes). The bounds on the decoding error probability of an ML decoder are compared to computer simulations of iterative decoding. The new bounds show a remarkable improvement over the union bound and some other previously reported bounds for independent parallel channels. This improvement is exemplified for relatively short block lengths, and it is pronounced when the block length is increased. In the asymptotic case where we let the block length tend to infinity, inner bounds on the attainable channel regions of modern coding techniques under ML decoding are obtained, based solely on the asymptotic growth rates of the average distance spectra of these code ensembles.

Index Terms

Accumulate-repeat-accumulate codes, distance spectrum, input-output weight enumerator (IOWE), iterative decoding, linear codes, maximum-likelihood (ML) decoding, memoryless binary-input output-symmetric (MBIOS) channels, parallel channels.

I. INTRODUCTION

The transmission of coded communication systems is widely modeled to take place over a set of parallel channels. This situation can be modelled as having a set of independent parallel channels, where the transmitted codeword is partitioned into disjoint sets, and the symbols within each set are transmitted over one of these channels. Some examples in which this scenario may be used include block-fading channels (for performance bounds of coded communication systems over block-fading channels, see, e.g., [16], [38]), rate-compatible puncturing of turbo-like codes (see, e.g., [17], [35]), incremental redundancy retransmission schemes, cooperative coding, multi-carrier signaling (for performance bounds of coded orthogonal-frequency division multiplexing (OFDM) systems, see e.g., [39]), and other applications.

Tight analytical bounds serve as a potent tool for assessing the performance of modern error-correction schemes, both for the case of finite block length and in the asymptotic case where the block length tends to infinity. In the setting of a single communication channel and by letting the block length tend to infinity, these bounds are applied in order to obtain a noise threshold which indicates the worst channel conditions necessary for reliable communication. When generalizing the bounds to the scenario of independent parallel channels, this threshold is transformed into a multi-dimensional barrier within the space of the joint parallel-channel transition probabilities,

dividing the space into channel regions where reliable communication is available and where it is not. One of the most widespread upper bounds for a single channel is the union bound, which is easily applied to the analysis of many communication systems. Its main drawback is that for codes of large enough block lengths, it is useless for rates exceeding the channel cutoff rate. Modern communication systems are required to operate well beyond this rate. Therefore, tighter upper bounds are required in order to assess the performance of such systems. When considering upper bounds for a single channel or independent parallel channels, it is desirable to have the bound expressible in terms of basic features of the code, such as the distance spectrum. Sometimes the distance spectrum cannot be evaluated for a specific code, but rather, an ensemble average can be obtained. Consequently, another desirable feature of any upper bound is to be applicable to ensembles of codes as well as to particular codes.

Tight upper bounds on the ML decoding error probability which can be applied to specific codes as well as structured ensembles of codes and which depend on the distance spectrum of the code (or ensemble) date back to Gallager [14]. Other examples of tight upper bounds include the generalized second version of the Duman-Salehi bound (often termed as the DS2 bound) [12], [36], the tangential sphere bound [30], the Shulman and Feder bound [37], and others. In this respect, it was shown by Sason and Shamai [36] that many reported upper bounds are special cases of the DS2 bound, including the 1961 Gallager bound [14]. For a comprehensive monograph on performance bounds of linear codes under ML decoding, the reader is referred to [33].

In his thesis [13], Ebert considered the problem of communicating over parallel discrete-time channels, disturbed by arbitrary and independent additive Gaussian noises, where a total power constraint is imposed upon the channel inputs. He found explicit upper and lower bounds on the ML decoding error probability, which decrease exponentially with the block length. The exponents of the upper and lower bounds coincide at zero rate and at all rates between the critical rate (R_{crit}) and capacity. The results were also shown to be applicable to colored Gaussian noise channels with an average power constraint on the channel. However, this work refers only to random codes and does not apply to specific codes or structured ensembles of codes.

The main difficulty which arises in the analysis of specific codes transmitted over parallel channels stems from the inherent asymmetry of the parallel-channel setting, which poses a difficulty for the analysis, as different symbols of the codeword suffer varying degrees of degradation through the different parallel channels. This difficulty was circumvented in [23] by introducing a random mapper, i.e., a device which randomly and independently assigns symbols to the different channels according to a certain a-priori probability distribution. As a result of this randomization, Liu et al. [23] derived upper bounds on the ML decoding error probability which depend solely on the weight enumerator of the overall code, instead of a specific split weight enumerator which follows from the partitioning of a codeword into several subsets of bits and the transmission of each subset over a different channel. The analysis in [23] modifies the 1961 Gallager bound from [14, Chapter 3] and adapts this bounding technique for communication over parallel channels. However, the results presented in [23] rely on *special cases* of the 1961 Gallager bound for parallel channels and not on the optimized version of this bound. These special cases include a generalization of the union-Bhattacharyya bound, the Shulman-Feder bound [37], simplified sphere bound [10], and a combination of the two former bounds. Our motivation is two-fold: First, the 1961 Gallager bound for parallel channels can be improved by choosing optimized parameters and tilting measures. Secondly, the DS2 bound ([11], [33], [36]) can be generalized to parallel channels.

Using the approach of the random mapper by Liu et al. [23], we derive a parallel-channel generalization of the DS2 bound [11], [33], [36] via two different approaches. The comparison between these bounds yields that for random codes, one of the bounds is tighter than the other and achieves the channel capacity, while for a general ensemble, neither of these bounds is necessarily tighter. For the scenario where transmission takes place over parallel independent channels, we re-examine the inter-connections between the DS2 bound and the 1961 Gallager bound which were previously reported for the single channel-case [10], [33], [36]. In this respect, it is shown in this paper that one of the versions of the generalized DS2 bound is tighter than the corresponding generalization of the 1961 Gallager bound while the other is not necessarily tighter.

In the asymptotic case where we let the block length tend to infinity and the transmission takes place over a set of independent parallel channels, we obtain inner bounds on the boundary of the channel regions which ensure reliable communication under ML decoding; these results improve remarkably on those recently reported in [23]. The tightness of these bounds for independent parallel channels is exemplified for structured ensembles of turbo-like codes, and the boundary of the improved attainable channel regions is compared with previously reported regions for Gaussian parallel channels. It shows significant improvement due the optimization of the tilting measures which

are involved in the computation of the generalized DS2 and 1961 Gallager bounds for parallel channels.

The remainder of the paper is organized as follows. The system model is presented in Section II, as well as preliminary material related to our discussion. In Section III, we generalize the DS2 bound to the case of independent parallel channels. Section IV presents the 1961 Gallager bound from [23], and considers its connection to the generalized DS2 bound, along with the optimization of its tilting measures. Section V presents some special cases of these upper bounds which are obtained as particular cases of the generalized bounds in Sections III and IV. Attainable channel regions are derived in Section VI. Inner bounds on attainable channel regions for various ensembles of turbo-like codes and performance bounds for moderate block lengths are exemplified in Section VII. Finally, Section VIII concludes the paper.

II. PRELIMINARIES

In this section, we state the assumptions on which our analysis is based. We also introduce notation and preliminary material related to the performance analysis of binary linear codes whose transmission takes place over a set of independent parallel channels.

A. System Model

We consider the case where the communication model consists of a parallel concatenation of J statistically independent MBIOS channels, as shown in Fig. 1.
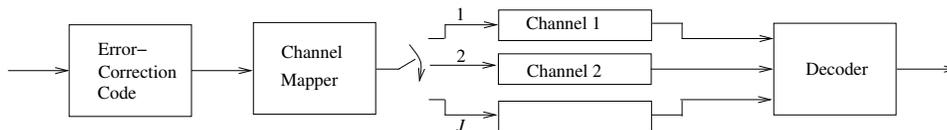

Fig. 1. System model of parallel channels. A random mapper is assumed where every bit is assigned to one of the J channels; a bit is assigned to the jth channel independently of the other bits and with probability α_j (where $\sum_{j=1}^{J} \alpha_j = 1$).

Using a linear error-correcting code $\mathcal{C}$ of size $M = 2^k$, the encoder selects a codeword $\underline{x}^m$ ($m = 0, 1, \ldots, M-1$) to be transmitted, where all codewords are assumed to be selected with equal probability ($\frac{1}{M}$). Each codeword consists of n symbols and the coding rate is defined as $R \triangleq \frac{\log_2 M}{n} = \frac{k}{n}$; this setting is referred to as using an (n, k) code. The channel mapper selects for each coded symbol one of J channels through which it is transmitted. The j-th channel component is characterized by a transition probability $p(y|x; j)$. The considered model assumes that the channel encoder performs its operation without prior knowledge of the specific mapping of the bits to the parallel channels. While in reality, the choice of the specific mapping is subject to the levels of importance of different coded bits, the considered model assumes for the sake of analysis that this mapping is random and independent of the coded bits. This assumption enables to average over all possible mappings, though suitable choices of mappings for the coded bits are expected to perform better than the average.

The received vector $\underline{y}$ is maximum-likelihood (ML) decoded at the receiver, where the specific channel mapper is assumed to be known. While this broad setting gives rise to very general coding, mapping and decoding schemes, we will focus on the case where the input alphabet is binary, i.e., $x \in \{-1, 1\}$ (where zero and one are mapped to $+1$ and -1, respectively). The output alphabet is real, and may be either finite or continuous. By its definition, the mapping device divides the set of indices $\{1, \ldots, n\}$ into J disjoint subsets $\mathcal{I}(j)$ for $j = 1, \ldots, J$, and transmits all the bits whose indices are included in the subset $\mathcal{I}(j)$ through the j-th channel. We will see in the next section that for a fixed channel mapping device (i.e., for given sets $\mathcal{I}(j)$), the problem of upper-bounding the ML decoding error probability is exceedingly difficult. In order to circumvent this difficulty, a probabilistic mapping device was introduced in [23] which uses a random assignment of the bits to the J parallel channels; this random mapper takes a symbol and assigns it to channel j with probability α_j. This assignment is independent of that of other symbols, and by definition, the equality $\sum_{j=1}^{J} \alpha_j = 1$ follows. This approach enables the derivation of an upper bound in the setting of parallel channels (see [23]) which is averaged over all possible channel assignments, and the bound is calculated in terms of the distance spectrum of the code (or ensemble). Another benefit of the random mapping approach is that it naturally accommodates for practical settings where one is faced with parallel channels having different capacities.

B. Capacity Limit and Cutoff Rate of Parallel MBIOS Channels

We consider here the capacity and cutoff rate of independent parallel MBIOS channels. These information-theoretic quantities serve as a benchmark for assessing the gap under optimal ML decoding between the achievable channel regions for various ensembles of codes and the capacity region. It is also useful for providing a quantitative measure for the asymptotic performance of various ensembles.

1) Cutoff Rate: The cutoff rate of an MBIOS channel is given by

$$R_0 = 1 - \log_2(1 + \gamma) \tag{1}$$

where γ is the Bhattacharyya constant, i.e.,

$$\gamma \triangleq \sum_y \sqrt{p(y|0)p(y|1)}. \tag{2}$$

Clearly, for continuous-output channels, the sum in the RHS of (2) is replaced by an integral.

For parallel MBIOS channels where every bit is assumed to be independently and randomly assigned to one of J channels with a-priori probability α_j (where $\sum_{j=1}^{J} \alpha_j = 1$), the Bhattacharyya constant of the resulting channel is equal to the weighted sum of the Bhattacharyya constants of these individual channels, i.e.,

$$\gamma = \sum_{J=1}^{J} \left\{ \alpha_j \sum_y \sqrt{p(y|0; j)p(y|1; j)} \right\}. \tag{3}$$

Consider a set of J parallel binary-input AWGN channels characterized by the transition probabilities

$$
\begin{aligned}
p(y|0; j) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{(y+\sqrt{2\nu_j})^2}{2}} \\
p(y|1; j) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-\sqrt{2\nu_j})^2}{2}}
\end{aligned}
\tag{4}
$$

$$-\infty < y < \infty, \ j = 1, \ldots, J$$

where

$$\nu_j \triangleq R \left(\frac{E_b}{N_0} \right)_j \tag{5}$$

and $\left(\frac{E_b}{N_0} \right)_j$ is the energy per information bit to the one-sided spectral noise density of the j-th channel. In this case, the Bhattacharyya constant is given by

$$\gamma = \sum_{j=1}^{J} \alpha_j e^{-\nu_j} \tag{6}$$

where ν_j is introduced in (5). From (1) and (6), the cutoff rate of J parallel binary-input AWGN channels is given by

$$R_0 = 1 - \log_2 \left(1 + \sum_{j=1}^{J} \alpha_j e^{-R\left(\frac{E_b}{N_0} \right)_j} \right) \text{ bits per channel use.} \tag{7}$$

Consider the case of $J = 2$ parallel binary-input AWGN channels. Given the value of $\left(\frac{E_b}{N_0} \right)_1$, and the code rate R (in bits per channel use), it is possible to calculate the value of $\left(\frac{E_b}{N_0} \right)_2$ of the second channel which corresponds to the cutoff rate. To this end, we set R_0 in the LHS of (7) to R. Solving this equation gives

$$\left(\frac{E_b}{N_0} \right)_2 = -\frac{1}{R} \ln \left(\frac{2^{1-R} - 1 - \alpha_1 e^{-R\left(\frac{E_b}{N_0} \right)_1}}{\alpha_2} \right). \tag{8}$$

2) Capacity Limit: Let C_j designate the capacity (in bits per channel use) of the j-th MBIOS channel. Clearly, by symmetry considerations, the capacity-achieving input distribution for all these channels is $\underline{q} = \left(\frac{1}{2}, \frac{1}{2}\right)$. The capacity of the J parallel channels where each bit is randomly and independently assigned to the j-th channel with probability α_j is therefore given by

$$C = \sum_{j=1}^{J} \alpha_j C_j. \tag{9}$$

For the case of J parallel binary-input AWGN channels

$$C_j = 1 - \frac{1}{\sqrt{2\pi}\ln(2)} \int_{-\infty}^{\infty} e^{-\frac{(y-\beta_j)^2}{2}} \ln\left(1 + e^{-2\beta_j y}\right) dy \quad \text{bits per channel use} \tag{10}$$

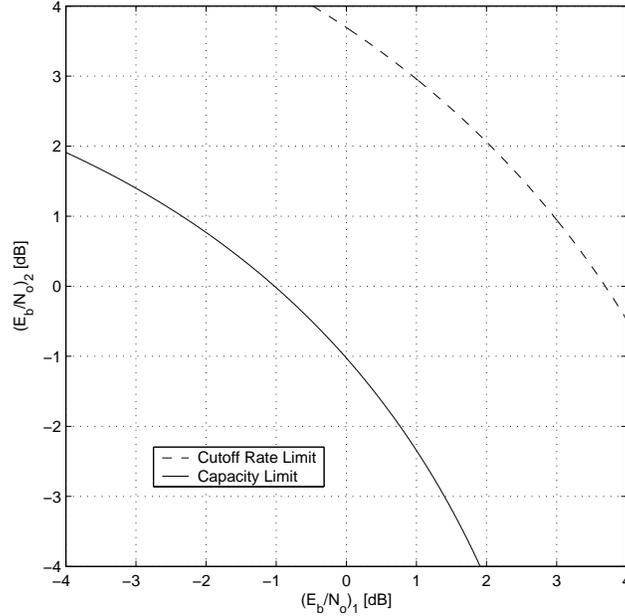where $\beta_j \triangleq \sqrt{2\nu_j}$ and ν_j is introduced in (5).

Fig. 2. Attainable channel regions for two parallel binary-input AWGN channels, as determined by the cutoff rate and the capacity limit, referring to a code rate of one-third bits per channel use. It is assumed that each bit is randomly and independently assigned to one of these channels with equal probability (i.e., $\alpha_1 = \alpha_2 = \frac{1}{2}$).

In order to simplify the numerical computation of the capacity, one can express each integral in (10) as a sum of two integrals from 0 to ∞, and use the power series expansion of the logarithmic function; this gives an infinite power series with alternating signs. Using the Euler transform to expedite the convergence rate of these infinite sums, gives the following alternative expression:

$$C_j = 1 - \frac{1}{\ln(2)} \left[\frac{2\beta e^{-\frac{\beta_j^2}{2}}}{\sqrt{2\pi}} - (2\beta_j^2 - 1)Q(\beta_j) + \sum_{k=0}^{\infty} \frac{(-1)^k \cdot \Delta^k a_0(j)}{2^{k+1}} \right], \quad j = 1, \ldots, J \tag{11}$$

where

$$\Delta^k a_0(j) \triangleq \frac{1}{2} e^{-\frac{\beta_j^2}{2}} \sum_{m=0}^{k} \left\{ \frac{(-1)^m}{(k-m+1)(k-m+2)} \binom{k}{m} \text{erfcx}\left(\frac{(2k-2m+3)\beta_j}{\sqrt{2}}\right) \right\}$$

and

$$\text{erfcx}(x) \triangleq 2e^{x^2} Q(\sqrt{2}x)$$

(note that $\text{erfcx}(x) \approx \frac{1}{\sqrt{\pi}} \cdot \frac{1}{x}$ for large values of x). The infinite sum in (11) converges exponentially fast with k, and the summation of its first 30 terms gives very accurate results irrespectively of the value of β_j.

Consider again the case of $J = 2$ parallel binary-input AWGN channels. Given the value of $\left(\frac{E_b}{N_0}\right)_1$, and the code rate R (in bits per channel use), (9) and (10) enable one to calculate the value of $\left(\frac{E_b}{N_0}\right)_2$ for the second channel, referring to the capacity limitation. To this end, one needs to set C in the LHS of (9) to the code rate R, and find the resulting value of $\left(\frac{E_b}{N_0}\right)_2$ which corresponds to the capacity limit. The boundary of the capacity region is represented by the continuous curve in Fig. 2 for $R = \frac{1}{3}$ bits per channel use; it is compared to the dashed curve in this figure which represents the boundary of the attainable channel region referring to the cutoff-rate limit (see Eq. (8)).

C. Distance Properties of Ensembles of Turbo-Like Codes

In this paper, we exemplify our numerical results by considering several ensembles of binary linear codes. Due to the outstanding performance of turbo-like codes, our focus is mainly on such ensembles, where we also consider as a reference the ensemble of fully random block codes which achieves capacity under ML decoding. The other ensembles considered in this paper include turbo codes, repeat-accumulate codes and some recent variations.

Bounds on the ML decoding error probability are often based on the distance properties of the considered codes or ensembles (see, e.g., [33] and references therein). The distance spectra and their asymptotic growth rates for various turbo-like ensembles were studied in the literature, e.g., for ensembles of uniformly interleaved repeat-accumulate codes and variations [2], [9], [18], ensembles of uniformly interleaved turbo codes [4], [5], [25], [34], and ensembles of regular and irregular LDPC codes [6], [8], [14], [24]. In this subsection, we briefly present the distance properties of some turbo-like ensembles considered in this paper.

Let us denote by $[\mathcal{C}(n)]$ an ensemble of codes of length n. We will also consider a *sequence of ensembles* $[\mathcal{C}(n_1)], [\mathcal{C}(n_2)], \ldots$ where all these ensembles possess a common rate R. For a given (n, k) linear code $\mathcal{C}$, let $A_h^{\mathcal{C}}$ (or simply A_h) denote the distance spectrum, i.e., the number of codewords of Hamming weight h. For a set of codes $[\mathcal{C}(n)]$, we define the *average distance spectrum* as

$$A_h^{[\mathcal{C}(n)]} \triangleq \frac{1}{|[\mathcal{C}(n)]|} \sum_{\mathcal{C} \in [\mathcal{C}(n)]} A_h^{\mathcal{C}} . \tag{12}$$

Let $\Psi_n \triangleq \{\delta : \delta = \frac{h}{n} \text{ for } h = 1, \ldots, n\} = \{\frac{1}{n}, \frac{2}{n}, \ldots, 1\}$ denote the set of *normalized distances*, then the *normalized exponent of the distance spectrum w.r.t. the block length* is defined as

$$r^{\mathcal{C}}(\delta) \triangleq \frac{\ln A_h^{\mathcal{C}}}{n} \quad , \quad r^{[\mathcal{C}(n)]}(\delta) \triangleq \frac{\ln A_h^{[\mathcal{C}(n)]}}{n}. \tag{13}$$

The motivation for this definition lies in our interest to consider the asymptotic case where $n \to \infty$. In this case we define the *asymptotic exponent of the distance spectrum* as

$$r^{[\mathcal{C}]}(\delta) \triangleq \lim_{n \to \infty} r^{[\mathcal{C}(n)]}(\delta) . \tag{14}$$

The input-output weight enumerator (IOWE) of a linear block code is given by a sequence $\{A_{w,h}\}$ designating the number of codewords of Hamming weight h which are encoded by information bits whose Hamming weight is w. Referring to ensembles, one considers the average IOWE and distance spectrum over the ensemble. The distance spectrum and IOWE of linear block codes are useful for the analysis of the block and bit error probabilities, respectively, under ML decoding.

As a reference to all ensembles, we will consider the ensemble of fully random block codes which is capacity-achieving under ML decoding (or 'typical pairs') decoding.

The ensemble of fully random binary block codes: Consider the ensemble of binary random codes $[\mathcal{RB}]$, where the set $[\mathcal{RB}(n, R)]$ consists of all binary codes of length n and rate R. For this ensemble, the following well-known equalities hold:

$$
\begin{aligned}
A_h^{[\mathcal{RB}(n,R)]} &= \binom{n}{h} 2^{-n(1-R)} \\
r^{[\mathcal{RB}(n,R)]}(\delta) &= \frac{\ln \binom{n}{h}}{n} - (1-R)\ln 2 \\
r^{[\mathcal{RB}(R)]}(\delta) &= H(\delta) - (1-R)\ln 2
\end{aligned}
\tag{15}
$$

where $H(x) \triangleq -x \ln(x) - (1-x) \ln(1-x)$ designates the binary entropy function to the natural base.

Non-systematic repeat-accumulate codes: The ensemble of uniformly interleaved and non-systematic repeat-accumulate (NSRA) codes [9] is defined as follows. The information block of length N is repeated q times by the encoder. The bits are then uniformly reordered by an interleaver of size qN, and, finally, encoded by a rate-1 differential encoder (accumulator), i.e., a truncated rate-1 recursive convolutional encoder with a transfer function $1/(1+D)$. The ensemble [NSRA(N,q)] is defined to be the set of $\frac{(qN)!}{(q!)^N N!}$ RA different codes when considering the different possible permutations of the interleaver.[1] The (average) IOWE of the ensemble of uniformly interleaved RA codes $\mathcal{RA}_q(N)$ was originally derived in [9, Section 5], and it is given by

$$A_{w,h}^{\text{NSRA}(N,q)} = \frac{\binom{N}{w}\binom{qN-h}{\lfloor \frac{qw}{2} \rfloor}\binom{h-1}{\lceil \frac{qw}{2} \rceil - 1}}{\binom{qN}{qw}}. \tag{16}$$

The average distance spectrum of the ensemble is therefore given by

$$A_h^{\text{NSRA}(N,q)} = \sum_{w=1}^{\min(N,\lfloor \frac{2h}{q} \rfloor)} \frac{\binom{N}{w}\binom{qN-h}{\lfloor \frac{qw}{2} \rfloor}\binom{h-1}{\lceil \frac{qw}{2} \rceil - 1}}{\binom{qN}{qw}}, \quad \left\lceil \frac{q}{2} \right\rceil \le h \le qN - \left\lfloor \frac{q}{2} \right\rfloor$$

where $A_h^{\text{NSRA}(N,q)} = 0$ for $1 \le h < \lceil \frac{q}{2} \rceil$, and $A_0^{\text{NSRA}(N,q)} = 1$ since the all-zero vector is always a codeword of a linear code. The asymptotic exponent of the distance spectrum of this ensemble is given by (see [19])

$$r^{[\text{NSRA}(q)]}(\delta) \triangleq \lim_{N \to \infty} r^{[\text{NSRA}(N,q)]}(\delta)$$
$$= \max_{0 \le u \le \min(2\delta, 2-2\delta)} \left\{ -\left(1 - \frac{1}{q}\right) H(u) + (1-\delta) H\left(\frac{u}{2(1-\delta)}\right) + \delta H\left(\frac{u}{2\delta}\right) \right\}. \tag{17}$$
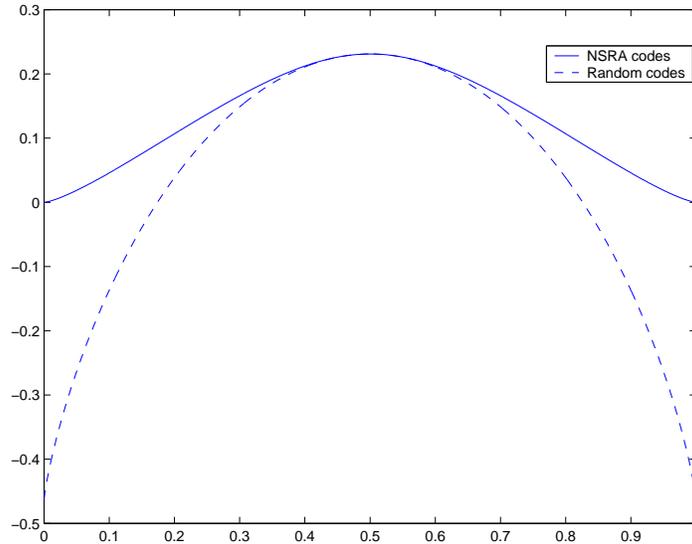

Fig. 3. Plot of the asymptotic exponent of the distance spectra for the ensemble of fully random block codes, and the ensemble of uniformly interleaved and non-systematic repeat-accumulate (NSRA) codes of rate $\frac{1}{3}$ bits per channel use. The curves are depicted as a function of the normalized Hamming weight (δ), and their calculations are based on (15) and (17).

The IOWEs and distance spectra of various ensembles of irregular repeat-accumulate (IRA) and accumulate-repeat-accumulate (ARA) codes are derived in [2], [18]. The reader is referred to Fig. 3, for a comparison between the asymptotic exponents (i.e., growth rates) of the distance spectra for the ensemble of fully random block codes and the ensemble of uniformly interleaved NSRA codes where both ensembles are assumed to be of rate one-third

[1]There are $(qN)!$ ways to place qN bits. However, permuting the q repetitions of any of the N information bits does not affect the result of the interleaving, so there are $\frac{(qN)!}{(q!)^N}$ possible ways for the interleaving. Strictly speaking, by permuting the N information bits, the vector space of the code does not change, which then yields that there are $\frac{(qN)!}{(q!)^N N!}$ distinct RA codes of dimension k and number of repetitions q.

bit per channel use. One can observe from Fig. 3 that the two curves deviate considerably at low and high values of the normalized Hamming weight, while there is a good match between these two curves for intermediate values of the normalized Hamming weight.

D. The DS2 Bound for a Single MBIOS Channel

The bounding technique of Duman and Salehi [11], [12] originates from the 1965 Gallager bound [15] which states that the conditional ML decoding error probability $P_{e|m}$ given that a codeword $\underline{x}^m$ (of block length n) is transmitted is upper-bounded by

$$P_{e|m} \leq \sum_{\underline{y}} p_n\left(\underline{y}|\underline{x}^m\right) \left(\sum_{m' \neq m} \left(\frac{p_n(\underline{y}|\underline{x}^{m'})}{p_n(\underline{y}|\underline{x}^m)}\right)^\lambda\right)^\rho \quad \lambda, \rho \geq 0 \tag{18}$$

where $p_n(\underline{y}|\underline{x})$ designates the conditional *pdf* of the communication channel to obtain an n-length sequence $\underline{y}$ at the channel output, given the n-length input sequence $\underline{x}$.

Unfortunately, this upper bound is not calculable in terms of the distance spectrum of the code ensemble, except for the ensembles of fully random block codes and orthogonal codes transmitted over a memoryless channel, and the special case where $\rho = 1, \lambda = 0.5$ in which the bound reduces to the union-Bhattacharyya bound. With the intention of alleviating the difficulty of calculating the bound for specific codes and ensembles, we introduce the function $\Psi_n^{(m)}(\underline{y})$ which is an arbitrary probability tilting measure. This function may depend in general on the index m of the transmitted codeword [36], and is a non-negative function which satisfies the equality $\int_{\underline{y}} \Psi_n^{(m)}(\underline{y}) \, d\underline{y} = 1$. The upper bound in (18) can be rewritten in the following equivalent form:

$$P_{e|m} \leq \sum_{\underline{y}} \Psi_n^{(m)}(\underline{y}) \left(\Psi_n^{(m)}(\underline{y})^{-\frac{1}{\rho}} p_n\left(\underline{y}|\underline{x}^m\right)^{\frac{1}{\rho}} \sum_{m' \neq m} \left(\frac{p_n(\underline{y}|\underline{x}^{m'})}{p_n(\underline{y}|\underline{x}^m)}\right)^\lambda\right)^\rho \quad \lambda, \rho \geq 0. \tag{19}$$

Recalling that $\Psi_n^{(m)}$ is a probability measure, we invoke Jensen's inequality in (19) which gives

$$P_{e|m} \leq \left(\sum_{m' \neq m} \sum_{\underline{y}} \Psi_n^{(m)}(\underline{y})^{1-\frac{1}{\rho}} p_n(\underline{y}|\underline{x}^m)^{\frac{1}{\rho}} \left(\frac{p_n(\underline{y}|\underline{x}^{m'})}{p_n(\underline{y}|\underline{x}^m)}\right)^\lambda\right)^\rho, \quad \begin{matrix} 0 \leq \rho \leq 1 \\ \lambda \geq 0 \end{matrix} \tag{20}$$

which is the DS2 bound. This expression can be simplified (see, e.g., [36]) for the case of a single memoryless channel where

$$p_n(\underline{y}|\underline{x}) = \prod_{i=1}^n p(y_i|x_i).$$

Let us consider probability tilting measures $\Psi_n^{(m)}(\underline{y})$ which can be factorized into the form

$$\Psi_n^{(m)}(\underline{y}) = \prod_{i=1}^n \psi^{(m)}(y_i)$$

recalling that the function $\psi^{(m)}$ may depend on the transmitted codeword $\underline{x}^m$. In this case, the bound in (20) is calculable in terms of the distance spectrum of the code, thus not requiring the fine details of the code structure.

Let $\mathcal{C}$ be a binary linear block code whose length is n, and let its distance spectrum be given by $\{A_h\}_{h=0}^n$. Consider the case where the transmission takes place over an MBIOS channel. By partitioning the code into subcodes of constant Hamming weights, let $\mathcal{C}_h$ be the set which includes all the codewords of $\mathcal{C}$ with Hamming weight h and the all-zero codeword. Note that this forms a partitioning of a linear code into subcodes which are in general non-linear. We apply the DS2 bound on the conditional ML decoding error probability (given the all-zero codeword is transmitted), and finally use the union bound w.r.t. the subcodes $\{\mathcal{C}_h\}$ in order to obtain an upper bound on the ML decoding error probability of the code $\mathcal{C}$. Referring to the constant Hamming weight subcode $\mathcal{C}_h$, the bound (20) gives

$$P_{e|0}(h) \leq (A_h)^\rho \left\{\left(\sum_y \psi(y)^{1-\frac{1}{\rho}} p(y|0)^{\frac{1}{\rho}}\right)^{n-h} \left(\sum_y \psi(y)^{1-\frac{1}{\rho}} p(y|0)^{\frac{1-\lambda\rho}{\rho}} p(y|1)^\lambda\right)^h\right\}^\rho \quad \begin{matrix} 0 \leq \rho \leq 1 \\ \lambda \geq 0 \end{matrix}. \tag{21}$$

Clearly, for an MBIOS channel with continuous output, the sums in (21) are replaced by integrals. In order to obtain the tightest bound within this form, the probability tilting measure ψ and the parameters λ and ρ are optimized. The optimization of ψ is based on calculus of variations, and is independent of the distance spectrum (this will be proved later also for the case of independent parallel MBIOS channels).

Due to the symmetry of the channel and the linearity of the code $\mathcal{C}$, the decoding error probability of $\mathcal{C}$ is independent of the transmitted codeword. Since the code $\mathcal{C}$ is the union of the subcodes $\{\mathcal{C}_h\}$, the union bound provides an upper bound on the ML decoding error probability of $\mathcal{C}$ which is expressed as the sum of the conditional decoding error probabilities of the subcodes $\mathcal{C}_h$ given that the all-zero codeword is transmitted. Let $d_{\min}$ be the minimum distance of the code $\mathcal{C}$, and R be the rate of the code $\mathcal{C}$. Based on the linearity of the code, the geometry of the Voronoi regions (see [3]) gives the following expurgated union bound:

$$P_{\mathrm{e}} \leq \sum_{h=d_{\min}}^{n(1-R)} P_{\mathrm{e}|0}(h). \tag{22}$$

For the bit error probability, one may partition a binary linear block code $\mathcal{C}$ into subcodes w.r.t. the Hamming weights of the information bits and the code bits. Let $\mathcal{C}_{w,h}$ designate the subcode which includes the all-zero codeword and all the codeowrds of $\mathcal{C}$ whose Hamming weight is h and whose information bits have Hamming weight w. An upper bound on the bit error probability of the code $\mathcal{C}$ is performed by calculating the DS2 upper bound on the conditional bit error probability for each subcode $\mathcal{C}_{w,h}$ (given that the all-zero codeword is transmitted), and applying the union bound over all these subcodes. Note that the number of these subcodes is at most quadratic in the block length of the code, so taking the union bound w.r.t. these subcodes does not affect the asymptotic tightness of the overall bound. Let $\{A_{w,h}\}$ designate the IOWE of the code $\mathcal{C}$ whose block length and dimension are equal to n and k, respectively. The conditional DS2 bound on the bit error probability was demonstrated in [32], [33] to be identical to the DS2 bound on the block error probability, except that the distance spectrum of the code

$$A_h = \sum_{w=0}^{k} A_{w,h}, \quad h = 0, \ldots, n \tag{23}$$

appearing in the RHS of (21) is replaced by

$$A'_h \triangleq \sum_{w=0}^{k} \left(\frac{w}{k}\right) A_{w,h}, \quad h = 0, \ldots, n. \tag{24}$$

Since $A'_h \leq A_h$ then, as expected, the upper bound on the bit error probability is smaller than the upper bound on the block error probability.

Finally, note that the DS2 bound is also applicable to ensembles of linear codes. To this end, one simply needs to replace the distance spectrum or the IOWE of a code by the average quantities over this ensemble. This follows easily by invoking Jensen's inequality to the RHS of (21) which yields that $\mathbb{E}[(A_h)^\rho] \leq (\mathbb{E}[A_h])^\rho$ for $0 \leq \rho \leq 1$. The DS2 bound for a single channel is discussed in further details in [11], [32], [36] and the tutorial paper [33, Chapter 4].

III. GENERALIZED DS2 BOUNDS FOR INDEPENDENT PARALLEL CHANNELS

In this section, we generalize the DS2 bound to independent parallel MBIOS channels, and optimize the probability tilting measures in the generalized bound to obtain the tightest bound within this form. We will discuss two possible ways of generalizing the bound. These two versions of the bound are obtained via different way of looking on the set of parallel channels and their tightness is compared.

A. Generalizing the DS2 bound to Parallel Channels: First Approach

1) Derivation of the new bound in the first approach: Let us assume that the communication takes place over J statistically independent parallel channels where each one of the individual channels is memoryless binary-input output-symmetric (MBIOS) with antipodal signaling, i.e., $p(y|x = 1) = p(-y|x = -1)$. The essence of the approach discussed in this section is to start by considering the case of a specific channel assignment; the calculation then

proceeds by averaging the bound over all possible assignments. For a specific channel assignment, the assumption that all J channels are independent and MBIOS means that we factor the transition probability as

$$p_n\left(\underline{y}|\underline{x}^m\right) = \prod_{j=1}^{J}\prod_{i\in\mathcal{I}(j)} p(y_i|x_i^{(m)};j) \tag{25}$$

which we can plug into (20) to get a DS2 bound suitable for the case of parallel channels. In order to get a bound which depends on one-dimensional sums (or one-dimensional integrals), we impose a restriction on the tilting measure $\Psi_n^{(m)}(\cdot)$ in (20) so that it can be expressed as a J-fold product of one-dimensional probability tilting measures, i.e.,

$$\Psi_n^{(m)}(\underline{y}) = \prod_{j=1}^{J}\prod_{i\in\mathcal{I}(j)} \psi^{(m)}(y_i;j). \tag{26}$$

Considering a binary linear block code $\mathcal{C}$, the conditional decoding error probability does not depend on the transmitted codeword, so $P_{\mathrm{e}} \triangleq \frac{1}{M}\sum_{m=0}^{M-1} P_{\mathrm{e}|m} = P_{\mathrm{e}|0}$ where w.o.l.o.g., one can assume that the all-zero vector is the transmitted codeword.

The channel mapper for the J independent parallel channels is assumed to transmit the bits whose indices are included in the subset $\mathcal{I}(j)$ over the j-th channel where the subsets $\{\mathcal{I}(j)\}$ constitute a disjoint partitioning of the set of indices $\{1, 2, \ldots, n\}$.

Following the notation in [23], let $A_{h_1,h_2,\ldots,h_J}$ designate the *split weight enumerator* of the binary linear block code, defined as the number of codewords of Hamming weight h_j within the J disjoint subsets $\mathcal{I}(j)$ for $j = 1 \ldots J$. By substituting (25) and (26) in (20), we obtain

$$
\begin{aligned}
P_{\mathrm{e}} &= P_{\mathrm{e}|0} \\
&\leq \left\{ \sum_{h_1=0}^{|\mathcal{I}(1)|}\cdots\sum_{h_J=0}^{|\mathcal{I}(J)|}\sum_{\underline{y}} A_{h_1,h_2,\ldots,h_J}\prod_{j=1}^{J}\prod_{i\in\mathcal{I}(j)} \psi(y_i;j)^{1-\frac{1}{\rho}}p(y_i|0;j)^{\frac{1}{\rho}}\left(\frac{p(y_i|x_i;j)}{p(y_i|0;j)}\right)^{\lambda} \right\}^{\rho} \\
&= \left\{ \sum_{h_1=0}^{|\mathcal{I}(1)|}\cdots\sum_{h_J=0}^{|\mathcal{I}(J)|} A_{h_1,h_2,\ldots,h_J}\prod_{j=1}^{J}\prod_{i\in\mathcal{I}(j)}\sum_{y_i} \psi(y_i;j)^{1-\frac{1}{\rho}}p(y_i|0;j)^{\frac{1}{\rho}}\left(\frac{p(y_i|x_i;j)}{p(y_i|0;j)}\right)^{\lambda} \right\}^{\rho} \\
&= \left\{ \sum_{h_1=0}^{|\mathcal{I}(1)|}\cdots\sum_{h_J=0}^{|\mathcal{I}(J)|} A_{h_1,h_2,\ldots,h_J}\prod_{j=1}^{J}\left(\sum_{y} \psi(y;j)^{1-\frac{1}{\rho}}p(y|0;j)^{\frac{1-\lambda\rho}{\rho}}p(y|1;j)^{\lambda}\right)^{h_j} \right. \\
&\qquad\qquad \left. \prod_{j=1}^{J}\left(\sum_{y} \psi(y;j)^{1-\frac{1}{\rho}}p(y|0;j)^{\frac{1}{\rho}}\right)^{|\mathcal{I}(j)|-h_j} \right\}^{\rho} , \quad \begin{matrix} 0\leq\rho\leq 1 \\ \lambda\geq 0 \end{matrix} . \tag{27}
\end{aligned}
$$

We note that the bound in (27) is valid for a specific assignment of bits to the parallel channels. For structured codes or ensembles, the split weight enumerator is in general not available when considering specific assignments. As a result of this, we continue the derivation by using the random assignment approach. Let us designate $n_j \triangleq |\mathcal{I}(j)|$ to be the cardinality of the set $\mathcal{I}(j)$, so $E[n_j] = \alpha_j n$ is the expected number of bits assigned to channel no. j (where $j = 1, 2, \ldots, J$). Averaging (27) with respect to all possible channel assignments, we get the following bound on

the average ML decoding error probability:

$$P_e \leq \mathbf{E} \left\{ \sum_{h_1=0}^{n_1} \cdots \sum_{h_J=0}^{n_J} A_{h_1,h_2,\ldots,h_J} \prod_{j=1}^{J} \left(\sum_y \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1-\lambda\rho}{\rho}} p(y|1;j)^{\lambda} \right)^{h_j} \right.$$

$$\left. \prod_{j=1}^{J} \left(\sum_y \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1}{\rho}} \right)^{n_j-h_j} \right\}^{\rho}$$

$$= \sum_{\substack{n_j \geq 0 \\ \sum_j n_j = n}} \left\{ \sum_{h_1=0}^{n_1} \cdots \sum_{h_J=0}^{n_J} A_{h_1,h_2,\ldots,h_J} \prod_{j=1}^{J} \left(\sum_y \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1-\lambda\rho}{\rho}} p(y|1;j)^{\lambda} \right)^{h_j} \right.$$

$$\left. \prod_{j=1}^{J} \left(\sum_y \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1}{\rho}} \right)^{n_j-h_j} \right\}^{\rho} P_{\underline{N}}(\underline{n}) \qquad (28)$$

where $P_{\underline{N}}(\underline{n})$ designates the probability distribution of the discrete random vector $\underline{N} \triangleq (n_1,\ldots,n_J)$. Applying Jensen's inequality to the RHS of (28) and changing the order of summation give

$$P_e \leq \left\{ \sum_{\substack{n_j \geq 0 \\ \sum n_j = n}} \sum_{h=0}^{n} \sum_{\substack{h_1 \leq n_1,\ldots,h_J \leq n_J \\ h_1+\ldots+h_J=h}} A_{h_1,h_2,\ldots,h_J} P_{\underline{N}}(\underline{n}) \right.$$

$$\prod_{j=1}^{J} \left(\sum_y \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1-\lambda\rho}{\rho}} p(y|1;j)^{\lambda} \right)^{h_j}$$

$$\left. \prod_{j=1}^{J} \left(\sum_y \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1}{\rho}} \right)^{n_j-h_j} \right\}^{\rho}, \quad \begin{array}{c} 0 \leq \rho \leq 1 \\ \lambda \geq 0 \end{array} . \qquad (29)$$

Let the vector $\underline{H} = (h_1,\ldots,h_J)$ be the vector of partial Hamming weights referring to the bits transmitted over each channel (n_j bits are transmitted over channel no. j, so $0 \leq h_j \leq n_j$). Clearly, $\sum_{j=1}^{J} h_j = h$ is the overall Hamming weight of a codeword in $\mathcal{C}$. Due to the random assignment of the code bits to the parallel channels, we get

$$P_{\underline{N}}(\underline{n}) = \binom{n}{n_1, n_2, \ldots, n_J} \alpha_1^{n_1} \alpha_2^{n_2} \ldots \alpha_J^{n_J}$$

$$P_{\underline{H}|\underline{N}}(\underline{h}|\underline{n}) = \frac{\binom{h}{h_1,\ldots,h_J}\binom{n-h}{n_1-h_1,\ldots,n_J-h_J}}{\binom{n}{n_1,\ldots,n_J}}$$

$$A_{h_1,h_2,\ldots,h_J} P_{\underline{N}}(\underline{n})$$
$$= A_h \, P_{\underline{H}|\underline{N}}(\underline{h}|\underline{n}) \, P_{\underline{N}}(\underline{n})$$
$$= A_h \, \alpha_1^{n_1} \alpha_2^{n_2} \ldots \alpha_J^{n_J} \binom{h}{h_1,\ldots,h_J}\binom{n-h}{n_1-h_1,\ldots,n_J-h_J} \qquad (30)$$

and the substitution of (30) in (29) gives

$$
\begin{aligned}
P_{\mathrm{e}} \le \Bigg\{ & \sum_{\substack{n_j \ge 0 \\ \sum n_j = n}} \sum_{h=0}^{n} A_h \sum_{\substack{h_1 \le n_1, \dots, h_J \le n_J \\ h_1 + \dots + h_J = h}} \binom{h}{h_1, h_2, \dots, h_J} \\
& \binom{n-h}{n_1 - h_1, n_2 - h_2, \dots, n_J - h_J} \\
& \prod_{j=1}^{J} \left(\alpha_j \sum_y \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1-\lambda\rho}{\rho}} p(y|1;j)^{\lambda} \right)^{h_j} \\
& \prod_{j=1}^{J} \left(\alpha_j \sum_y \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1}{\rho}} \right)^{n_j - h_j} \Bigg\}^{\rho}.
\end{aligned}
$$

Let $k_j \triangleq n_j - h_j$ for $j = 1, 2, \dots, J$, then by changing the order of summation we obtain

$$
\begin{aligned}
P_{\mathrm{e}} \le \Bigg\{ & \sum_{h=0}^{n} A_h \sum_{\substack{h_1, \dots, h_J \ge 0 \\ h_1 + \dots + h_J = h}} \binom{h}{h_1, h_2, \dots, h_J} \prod_{j=1}^{J} \left(\alpha_j \sum_y \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1-\lambda\rho}{\rho}} p(y|1;j)^{\lambda} \right)^{h_j} \\
& \sum_{\substack{k_1, \dots, k_J \ge 0 \\ k_1 + \dots + k_J = n-h}} \binom{n-h}{k_1, k_2, \dots, k_J} \prod_{j=1}^{J} \left(\alpha_j \sum_y \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1}{\rho}} \right)^{k_j} \Bigg\}^{\rho}
\end{aligned}
$$

Since $\sum_{j=1}^{J} h_j = h$ and $\sum_{j=1}^{J} k_j = n - h$, the use of the multinomial formula gives

$$
\begin{aligned}
P_{\mathrm{e}} \le \Bigg\{ & \sum_{h=0}^{n} A_h \left(\sum_{j=1}^{J} \alpha_j \sum_y \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1-\lambda\rho}{\rho}} p(y|1;j)^{\lambda} \right)^{h} \\
& \left(\sum_{j=1}^{J} \alpha_j \sum_y \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1}{\rho}} \right)^{n-h} \Bigg\}^{\rho} \qquad
\begin{array}{c}
0 \le \rho \le 1 \\
\lambda \ge 0 \\
\sum_y \psi(y;j) = 1 \\
j = 1 \dots J
\end{array}
\end{aligned} \tag{31}
$$

which forms a possible generalization of the DS2 bound for independent parallel channels when averaging over all possible channel assignments. This result can be applied to specific codes as well as to structured ensembles for which the average distance spectrum $\overline{A_h}$ is known. In this case, the average ML decoding error probability $\overline{P_{\mathrm{e}}}$ is obtained by replacing A_h in (31) with the average distance spectrum $\overline{A_h}$ (this can be verified by noting that the function $f(t) = t^{\rho}$ is concave for $0 \le \rho \le 1$ and by invoking Jensen's inequality in (31)).

In the continuation of this section, we propose an equivalent version of the generalized DS2 bound for parallel channels where this equivalence follows the lines in [33], [36]. Rather than relying on a probability (i.e., normalized) tilting measure, the bound will be expressed in terms of an un-normalized tilting measure which is an arbitrary non-negative function. This version will be helpful later for the discussion on the connection between the DS2 bound and the 1961 Gallager bound for parallel channels, and also for the derivation of some particular cases of the DS2 bound. We begin by expressing the DS2 bound using the un-normalized tilting measure $G_n^{(m)}$ which is related to $\Psi_n^{(m)}$ by

$$
\Psi_n^{(m)}(\underline{y}) = \frac{G_n^{(m)}(\underline{y}) p_n(\underline{y}|\underline{x}^m)}{\sum_{\underline{y}'} G_n^{(m)}(\underline{y}') p_n(\underline{y}'|\underline{x}^m)} \ . \tag{32}
$$

Substituting (32) in (20) gives

$$
\begin{aligned}
P_{\mathrm{e}|m} \leq \quad & \left\{ \sum_{m' \neq m} \sum_{\underline{y}} G_n^{(m)}(\underline{y})^{1-\frac{1}{\rho}} p_n(\underline{y}|\underline{x}^m) \left(\frac{p_n(\underline{y}|\underline{x}^{m'})}{p_n(\underline{y}|\underline{x}^m)} \right)^{\lambda} \right\}^{\rho} \\
& \left(\sum_{\underline{y}} G_n^{(m)}(\underline{y}) p_n(\underline{y}|\underline{x}^m) \right)^{1-\rho} , \quad \begin{array}{c} 0 \leq \rho \leq 1 \\ \lambda \geq 0 \end{array} .
\end{aligned}
$$

As before, we assume that $G_n^{(m)}$ can be factored in the product form

$$
G_n^{(m)}(\underline{y}) = \prod_{j=1}^{J} \prod_{i \in \mathcal{I}(j)} g(y_i; j).
$$

Following the algebraic steps in (27)-(31) and averaging as before also over all the codebooks of the ensemble, we obtain the following upper bound on the ML decoding error probability:

$$
\begin{aligned}
P_{\mathrm{e}} = P_{\mathrm{e}|0} \leq \quad & \left\{ \sum_{h=0}^{n} A_h \left[\sum_{j=1}^{J} \alpha_j \left(\sum_{y} g(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{1-\lambda} p(y|1;j)^{\lambda} \right) \right. \right. \\
& \left(\sum_{y} g(y;j) p(y|0;j) \right)^{\frac{1-\rho}{\rho}} \bigg]^{h} \left[\sum_{j=1}^{J} \alpha_j \left(\sum_{y} g(y;j)^{1-\frac{1}{\rho}} p(y|0;j) \right) \right. \\
& \left. \left. \left(\sum_{y} g(y;j) p(y|0;j) \right)^{\frac{1-\rho}{\rho}} \right]^{n-h} \right\}^{\rho} , \quad \begin{array}{c} 0 \leq \rho \leq 1 \\ \lambda \geq 0 \end{array} . \quad (33)
\end{aligned}
$$

Note that the generalized DS2 bound as derived in this subsection is applied to the whole code (i.e., the optimization of the tilting measures refers to the whole code and is performed only once for each of the J channels). In the next subsection, we consider the partitioning of the code to constant Hamming weight subcodes, and then apply the union bound. For every such subcode, we rely on the conditional DS2 bound (given the all-zero codeword is transmitted), and optimize the J tilting measures; this optimization is performed for each of the subcodes separately. The total number of subcodes does not exceed the block length of the code (or ensemble), and hence the use of the union bound in this case does not degrade the related error exponent of the overall bound; moreover, the optimized tilting measures are tailored for each of the constant-Hamming weight subcodes, a process which can only improve the exponential behavior of the resulting bound.

2) Optimization of the Tilting Measures: In the following, we find optimized tilting measures $\{\psi(\cdot; j)\}_{j=1}^{J}$ which minimize the DS2 bound (31). The following calculation is a possible generalization of the analysis in [36] for a single channel to the considered case of an arbitrary number (J) of independent parallel MBIOS channels.

Let $\mathcal{C}$ be a binary linear block code of length n. Following the derivation in [23], [36], we partition the code $\mathcal{C}$ to constant Hamming weight subcodes $\{\mathcal{C}_h\}_{h=0}^{n}$, where $\mathcal{C}_h$ includes all the codewords of weight h $(h = 0, \ldots, n)$ as well as the all-zero codeword. Let $P_{\mathrm{e}|0}(h)$ denote the conditional block error probability of the subcode $\mathcal{C}_h$ under ML decoding, given that the all-zero codeword is transmitted. Based on the union bound, we get

$$
P_{\mathrm{e}} \leq \sum_{h=0}^{n} P_{\mathrm{e}|0}(h). \quad (34)
$$

As the code $\mathcal{C}$ is linear, $P_{\mathrm{e}|0}(h) = 0$ for $h = 0, 1, \ldots, d_{\min} - 1$ where $d_{\min}$ denotes the minimum distance of the code $\mathcal{C}$. The generalization of the DS2 bound in (31) gives the following upper bound on the conditional error

probability of the subcode $\mathcal{C}_h$:

$$P_{e|0}(h) \leq (A_h)^\rho \left\{ \left(\sum_{j=1}^{J} \alpha_j \sum_y \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1-\lambda\rho}{\rho}} p(y|1;j)^\lambda \right)^\delta \right.$$
$$\left. \left(\sum_{j=1}^{J} \alpha_j \sum_y \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1}{\rho}} \right)^{1-\delta} \right\}^{n\rho}, \quad \begin{array}{c} 0 \leq \rho \leq 1 \\ \lambda \geq 0 \end{array} \quad \delta \triangleq \frac{h}{n}. \tag{35}$$

Note that in this case, the set of probability tilting measures $\{\psi(\cdot;j)\}_{j=1}^{J}$ may also depend on the Hamming weight (h) of the subcode (or equivalently on δ). This is the result of performing the optimization on every individual constant-Hamming subcode instead of the whole code.

This generalization of the DS2 bound can be written equivalently in the exponential form

$$P_{e|0}(h) \leq e^{-nE_\delta^{\mathrm{DS2_1}}(\lambda,\rho,J,\{\alpha_j\})}, \quad 0 \leq \rho \leq 1, \quad \lambda \geq 0, \quad \delta \triangleq \frac{h}{n}. \tag{36}$$

where

$$E_\delta^{\mathrm{DS2_1}}(\lambda, \rho, J, \{\alpha_j\}) \triangleq -\rho r^{[\mathcal{C}]}(\delta) - \rho\delta \ln \left(\sum_{j=1}^{J} \alpha_j \sum_y \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1-\lambda\rho}{\rho}} p(y|1;j)^\lambda \right)$$
$$-\rho(1-\delta) \ln \left(\sum_{j=1}^{J} \alpha_j \sum_y \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1}{\rho}} \right) \tag{37}$$

and $r^{[\mathcal{C}]}(\delta)$ designates the normalized exponent of the distance spectrum as in (14).

Let

$$g_1(y;j) \triangleq p(y|0;j)^{\frac{1}{\rho}}, \quad g_2(y;j) \triangleq p(y|0;j)^{\frac{1}{\rho}} \left(\frac{p(y|1;j)}{p(y|0;j)} \right)^\lambda \tag{38}$$

then, for a given pair of λ and ρ (where $\lambda \geq 0$ and $0 \leq \rho \leq 1$), we need to minimize

$$\delta \ln \left(\sum_{j=1}^{J} \alpha_j \sum_y \psi(y;j)^{1-\frac{1}{\rho}} g_2(y;j) \right) + (1-\delta) \ln \left(\sum_{j=1}^{J} \alpha_j \sum_y \psi(y;j)^{1-\frac{1}{\rho}} g_1(y;j) \right)$$

over the set of non-negative functions $\psi(\cdot;j)$ satisfying the constraints

$$\sum_y \psi(y;j) = 1, \quad j = 1 \ldots J. \tag{39}$$

To this end, calculus of variations provides the following set of equations:

$$\psi(y;j)^{-\frac{1}{\rho}} \left(\frac{\alpha_j(1-\delta)(1-\frac{1}{\rho})g_1(y;j)}{\sum_y \sum_{j=1}^{J} \alpha_j \psi(y;j)^{1-\frac{1}{\rho}} g_1(y;j)} \right.$$
$$\left. + \frac{\alpha_j \delta(1-\frac{1}{\rho})g_2(y;j)}{\sum_y \sum_{j=1}^{J} \alpha_j \psi(y;j)^{1-\frac{1}{\rho}} g_2(y;j)} \right) + \xi_j = 0, \quad j = 1, \ldots, J \tag{40}$$

where ξ_j is a Lagrange multiplier. The solution of (40) is given in the following implicit form:

$$\psi(y;j) = \left(k_{1,j} g_1(y;j) + k_{2,j} g_2(y;j) \right)^\rho, \quad k_{1,j}, k_{2,j} \geq 0, \quad j = 1, \ldots, J$$

where

$$\frac{k_{2,j}}{k_{1,j}} = \frac{\delta}{1-\delta} \frac{\displaystyle\sum_{j=1}^{J} \sum_{y \in \mathcal{Y}} \alpha_j \psi(y;j)^{1-\frac{1}{\rho}} g_1(y;j)}{\displaystyle\sum_{j=1}^{J} \sum_{y \in \mathcal{Y}} \alpha_j \psi(y;j)^{1-\frac{1}{\rho}} g_2(y;j)}. \tag{41}$$

We note that $k \triangleq \frac{k_{2,j}}{k_{1,j}}$ in the RHS of (41) is *independent* of j. Thus, the substitution $\beta_j \triangleq k_{1,j}^\rho$ gives that the optimal tilting measures can be expressed as

$$
\begin{aligned}
\psi(y; j) &= \beta_j \big(g_1(y; j) + k g_2(y; j) \big)^\rho \\
&= \beta_j \, p(y|0; j) \left[1 + k \left(\frac{p(y|1; j)}{p(y|0; j)} \right)^\lambda \right]^\rho \quad y \in \mathcal{Y} \quad j = 1, \ldots, J.
\end{aligned}
\tag{42}
$$

By plugging (38) into (41) we obtain

$$
k = \frac{\delta}{1 - \delta} \frac{\displaystyle\sum_{j=1}^{J} \sum_{y \in \mathcal{Y}} \left\{ \alpha_j \beta_j^{1 - \frac{1}{\rho}} p(y|0; j) \left[1 + k \left(\frac{p(y|1; j)}{p(y|0; j)} \right)^\lambda \right]^{\rho-1} \right\}}{\displaystyle\sum_{j=1}^{J} \sum_{y \in \mathcal{Y}} \left\{ \alpha_j \beta_j^{1 - \frac{1}{\rho}} p(y|0; j) \left(\frac{p(y|1; j)}{p(y|0; j)} \right)^\lambda \left[1 + k \left(\frac{p(y|1; j)}{p(y|0; j)} \right)^\lambda \right]^{\rho-1} \right\}}
\tag{43}
$$

and from (38) and (39), β_j which is the appropriate factor normalizing the probability tilting measure $\psi(\cdot; j)$ in (42) is given by

$$
\beta_j = \left[\sum_y p(y|0; j) \left(1 + k \left(\frac{p(y|1; j)}{p(y|0; j)} \right)^\lambda \right)^\rho \right]^{-1}, \quad j = 1, \ldots, J.
\tag{44}
$$

Note that the implicit equation for k in (43) and the normalization coefficients in (44) provide a possible generalization of the results derived in [32, Appendix A] (where k is replaced there by α). The point here is that the value of $\frac{k_{2,j}}{k_{1,j}}$ in (41) is independent of j (where $j \in \{1, 2, \ldots, J\}$), a property which significantly simplifies the optimization process of the J tilting measures, and leads to the result in (42).

For the numerical calculation of the bound in (35) as a function of the normalized Hamming weight $\delta \triangleq \frac{h}{n}$, and for a fixed pair of λ and ρ (where $\lambda \geq 0$ and $0 \leq \rho \leq 1$), we find the optimized tilting measures in (42) by first assuming an initial vector $\overline{\beta^{(0)}} = (\beta_1, \ldots, \beta_J)$ and then iterating between (43) and (44) until we get a fixed point for these equations. For a fixed δ, we need to optimize numerically the bound in (36) w.r.t. the two parameters λ and ρ.

B. Generalizing the DS2 bound to Parallel Channels: Second Approach

1) Derivation of the new bound in a second approach: In this section we show a second way of generalizing the DS2 bound for independent parallel MBIOS channels. We begin by suggesting a system model equivalent to the one presented in Sec. II-A which we term the *channel side information at the receiver* (CSIR) model. Rather than viewing the set of component channels as parallel channels, we consider j (where $1 \leq j \leq J$) to be the internal state of a state-dependent channel $p(y|x; j)$ to which x is the input and y is the output. As in the parallel-channel model shown in Fig. 1, j is chosen at random for each transmitted symbol according to the a-priori probability distribution $\{\alpha_j\}$ from the finite alphabet $\{1, 2, \ldots, J\}$. Therefore, these two channel models are identical, except that we have to include the receiver's perfect knowledge of the channel state in the CSIR model. This is easily accomplished by viewing the internal state j as part of the output of the channel, i.e., the output is the pair $b \triangleq (y, j)$; the transition probability of this channel is thus denoted by $p_B(b|x)$. Since the channel and channel mapper both operate in a memoryless manner, the CSIR channel model is also memoryless. Finally, the transition probability $p_B(b|x)$ satisfies the relation

$$
p_B(b|x) = \alpha_j p(y|x; j)
\tag{45}
$$

because the channel state is independent of the input. If we define $-b \triangleq (-y, j)$, then we obtain from (45) and the symmetry of the transition probabilities $p(y|x; j)$ that $p_B(b|x) = p_B(-b| - x)$; thus, the CSIR model is also symmetric. In summary, the parallel-channel model presented in Sec. II-A is equivalent to an MBIOS channel with transition probability given in (45). We may thus use the DS2 bounding technique directly on the CSIR model; using this approach, the need to average over all channel mappings is circumvented.

Following this approach, we set the channel output to be $b = (y, j)$ and substitute (45) into (21) to get the upper bound

$$P_{\text{e}} \leq (A_h)^\rho \left(\sum_{j=1}^{J} \alpha_j^{\frac{1}{\rho}} \sum_y \psi(y; j)^{1 - \frac{1}{\rho}} p(y|0; j)^{\frac{1 - \lambda\rho}{\rho}} p(y|1; j)^\lambda \right)^h$$

$$\left(\sum_{j=1}^{J} \alpha_j^{\frac{1}{\rho}} \sum_y \psi(y; j)^{1 - \frac{1}{\rho}} p(y|0; j)^{\frac{1}{\rho}} \right)^{n-h} \qquad \begin{array}{c} 0 \leq \rho \leq 1 \\ \lambda \geq 0 \\ \sum_{y,j} \psi(y; j) = 1 \end{array} \tag{46}$$

As in the first approach (see (33)), this bound may also be expressed in terms of an un-normalized tilting measure, rather than a normalized (probability) measure. We will use this version later when we discuss special cases of this bound. The DS2 bound for parallel channels obtained using the second approach which is expressed using the un-normalized tilting measure is as follows:

$$P_{\text{e}} \leq (A_h)^\rho \left[\sum_{j=1}^{J} \alpha_j \sum_y g(y; j) p(y|0; j) \right]^{n(1-\rho)}$$

$$\left[\sum_{j=1}^{J} \alpha_j \left(\sum_y g(y; j)^{1 - \frac{1}{\rho}} p(y|0; j)^{1 - \lambda} p(y|1; j)^\lambda \right) \right]^{h\rho}$$

$$\left[\sum_{j=1}^{J} \alpha_j \left(\sum_y g(y; j)^{1 - \frac{1}{\rho}} p(y|0; j) \right) \right]^{(n-h)\rho} , \qquad \begin{array}{c} 0 \leq \rho \leq 1 \\ \lambda \geq 0 \end{array} . \tag{47}$$

We turn our attention to the derivation of optimized tilting measures for the generalized DS2 bound obtained using the second approach.

2) Optimization of the Tilting Measures: The optimization of tilting measures for the generalized DS2 bound in (46) obtained using the perfect CSIR model relies on this optimization for MBIOS channels. As in the first approach, the bound for a specific constant Hamming-weight subcode is expressed in exponential form

$$P_{\text{e}|0}(h) \leq e^{-n E_\delta^{\text{DS2}_2}(\lambda, \rho, J, \{\alpha_j\})}, \quad 0 \leq \rho \leq 1, \quad \lambda \geq 0, \quad \delta \triangleq \frac{h}{n} \tag{48}$$

where

$$E_\delta^{\text{DS2}_2}(\lambda, \rho, J, \{\alpha_j\}) \triangleq -\rho r^{[\mathcal{C}]}(\delta) - \rho\delta \ln \left(\sum_{j=1}^{J} \alpha_j^{\frac{1}{\rho}} \sum_y \psi(y; j)^{1 - \frac{1}{\rho}} p(y|0; j)^{\frac{1 - \lambda\rho}{\rho}} p(y|1; j)^\lambda \right)$$

$$- \rho(1 - \delta) \ln \left(\sum_{j=1}^{J} \alpha_j^{\frac{1}{\rho}} \sum_y \psi(y; j)^{1 - \frac{1}{\rho}} p(y|0; j)^{\frac{1}{\rho}} \right). \tag{49}$$

The optimized tilting measure should be chosen so as to maximize the exponent in (49). Since the perfect CSIR model is equivalent to an MBIOS channel, we can use the results of Sec. III-A.2 with $J = 1$; by substituting the transition probability from (45) into (42), we obtain that the optimal form of the tilting measure is given by

$$\psi(y; j) = \beta \alpha_j p(y|0; j) \left(1 + k \left(\frac{p(y|1; j)}{p(y|0; j)} \right)^\lambda \right)^\rho \tag{50}$$

where k is a parameter to be optimized and β is a normalizing constant given by

$$\beta = \left[\sum_{y,j} \alpha_j p(y|0; j) \left(1 + k \left(\frac{p(y|1; j)}{p(y|0; j)} \right)^\lambda \right)^\rho \right]^{-1} \tag{51}$$

C. Comparison Between the Two Generalized DS2 Bounds for Parallel Channels

Let us examine the two generalizations of the DS2 bound proposed in Sections III-A and III-B for the purpose of comparison. To this end, for constant weight subcodes of Hamming weight h (including the all-zero codeword), we write out the explicit expressions for the two bounds, including the optimal form of the tilting measures. By substituting (42) with the optimal value of k in (43), the bound in (31) obtained by the first approach reads

$$
\begin{aligned}
P_{\text{e}|0}(h) &\leq (A_h)^\rho \Bigg\{ \Bigg[\sum_{j=1}^{J} \alpha_j \sum_y \beta_j^{1-\frac{1}{\rho}} p(y|0;j) \left(\frac{p(y|1;j)}{p(y|0;j)} \right)^\lambda \left(1 + k_{\text{opt}}^{(1)} \left(\frac{p(y|1;j)}{p(y|0;j)} \right)^\lambda \right)^{\rho-1} \Bigg]^h \\
&\quad \Bigg[\sum_{j=1}^{J} \alpha_j \sum_y \beta_j^{1-\frac{1}{\rho}} p(y|0;j) \left(1 + k_{\text{opt}}^{(1)} \left(\frac{p(y|1;j)}{p(y|0;j)} \right)^\lambda \right)^{\rho-1} \Bigg]^{n-h} \Bigg\}^\rho .
\end{aligned}
\tag{52}
$$

In the same way, substituting (50) in (46) gives the bound obtained by using the second approach

$$
\begin{aligned}
P_{\text{e}|0}(h) &\leq (A_h)^\rho \Bigg\{ \Bigg[\sum_{j=1}^{J} \alpha_j \sum_y \beta^{1-\frac{1}{\rho}} p(y|0;j) \left(\frac{p(y|1;j)}{p(y|0;j)} \right)^\lambda \left(1 + k_{\text{opt}}^{(2)} \left(\frac{p(y|1;j)}{p(y|0;j)} \right)^\lambda \right)^{\rho-1} \Bigg]^h \\
&\quad \Bigg[\sum_{j=1}^{J} \alpha_j \sum_y \beta^{1-\frac{1}{\rho}} p(y|0;j) \left(1 + k_{\text{opt}}^{(2)} \left(\frac{p(y|1;j)}{p(y|0;j)} \right)^\lambda \right)^{\rho-1} \Bigg]^{n-h} \Bigg\}^\rho .
\end{aligned}
\tag{53}
$$

From these expressions one cannot conclusively deduce the superiority of one of the bounds over the other in general. However, in the *random coding setting*, it can be shown that the DS2 bound in Section III-B is tighter than the one in Section III-A. To this end, we show that the former bound attains the random coding exponent [15] while the latter does not.

The random coding exponent which corresponds to the MBIOS channel given by the perfect CSIR model, from which the second version in Section III-B is derived, gives the relation

$$
P_{\text{e}} \leq 2^{-n(E_0(\rho) - \rho R)} \quad 0 \leq \rho \leq 1
\tag{54}
$$

where

$$
\begin{aligned}
E_0(\rho) &= -\log_2 \left(\sum_b \left(\frac{1}{2} p_B(b|0)^{\frac{1}{1+\rho}} + \frac{1}{2} p_B(b|1)^{\frac{1}{1+\rho}} \right)^{1+\rho} \right) \\
&= -\log_2 \left(\sum_{j=1}^{J} \alpha_j \sum_y \left(\frac{1}{2} p(y|0;j)^{\frac{1}{1+\rho}} + \frac{1}{2} p(y|1;j)^{\frac{1}{1+\rho}} \right)^{1+\rho} \right)
\end{aligned}
\tag{55}
$$

We now turn to find the random coding exponent which stems from the use of the bound in Section III-B. We start with the bound in (47) which is expressed in terms of the un-normalized tilting measure. Consider the following choice for the un-normalized tilting measure

$$
g(y;j) = \left[\frac{1}{2} p(y|0;j)^{\frac{1}{1+\rho}} + \frac{1}{2} p(y|1;j)^{\frac{1}{1+\rho}} \right]^\rho p(y|0;j)^{-\frac{\rho}{1+\rho}} \quad , \quad j = 1, 2, \ldots, J
\tag{56}
$$

and the distance spectrum of the ensemble of random binary block codes of length n and rate R, given by

$$
A_h = 2^{-n(1-R)} \binom{n}{h}, \qquad h = 0, 1, \ldots, n.
\tag{57}
$$

Substituting (56) and (57) into (33) and setting $\lambda = \frac{1}{1+\rho}$ gives the bound

$$
P_{\text{e}} \leq 2^{nR\rho} \left\{ \sum_{j=1}^{J} \alpha_j \sum_y \left[\frac{1}{2} p(y|0;j)^{\frac{1}{1+\rho}} + \frac{1}{2} p(y|1;j)^{\frac{1}{1+\rho}} \right]^{1+\rho} \right\}^n
\tag{58}
$$

which coincides with the random coding bound in (54)-(55).

By substituting the tilting measure (56) in the bound in Section III-A (see (31)) we get the following error exponent, which appears instead of $E_0(\rho)$ in (55)

$$\tilde{E}_0(\rho) = -\log_2 \left(\sum_{j=1}^{J} \alpha_j \left[\left(\sum_y \frac{1}{2} p(y|0;j)^{\frac{1}{1+\rho}} + \frac{1}{2} p(y|1;j)^{\frac{1}{1+\rho}} \right)^{1+\rho} \right]^{\frac{1}{\rho}} \right)^{\rho}.$$

Using Jensen's inequality and the fact that $0 \leq \rho \leq 1$, it is easy to show that $\tilde{E}_0(\rho) \leq E_0(\rho)$, and we therefore conclude that the bound from Section III-B is tighter than the one in Section III-A in the random coding setting.

Discussion. When comparing the two versions of the bound, it should be noted that the two optimized forms of tilting measures as given in (42) and (50) are not identical. While these two forms of tilting measures exhibit the same functional behavior, the normalization conditions are slightly different, with J normalizing constants in the first version of the bound (see (44)) and one constant (see (51)) in the second version. This hints that either of these bounds in (52) and (53) is not uniformly tighter than the other for general codes or ensembles; this was also verified numerically by comparing the two bounds for some code ensembles. For random codes, we note that the tightness of the first version is hindered by the use of Jensen's inequality which is applied in the process of averaging over all possible channel assignments (see the move from (28) to (29)). This application of Jensen's inequality does not appear in the derivation of the second version of the DS2 bound, and may be the seed of the pitfall of the first version, when applied for random codes.

D. Statement of the Main Result Derived in Section III

The analysis in this section leads to the following theorem:

Theorem 1 (Generalized DS2 bounds for independent parallel MBIOS channels): Consider the transmission of binary linear block codes (or ensembles) over a set of J independent parallel MBIOS channels. Let the *pdf* of the jth MBIOS channel be given by $p(\cdot|0;j)$ where due to the symmetry of the binary-input channels $p(y|0;j) = p(-y|1;j)$. Assume that the coded bits are randomly and independently assigned to these channels, where each bit is transmitted over one of the J MBIOS channels. Let α_j be the a-priori probability of transmitting a bit over the jth channel $(j = 1, 2, \ldots, J)$, so that $\alpha_j \geq 0$ and $\sum_{j=1}^{J} \alpha_j = 1$. By partitioning the code into constant Hamming-weight subcodes, Eqs. (35) and (46) provide two possible upper bounds on the conditional ML decoding error probability for each of these subcodes, given that the all-zero codeword is transmitted, and (34) forms an upper bound on the block error probability of the whole code (or ensemble). For the bound in (35), the optimized set of probability tilting measures $\{\psi(\cdot;j)\}_{j=1}^{J}$ which attains the minimal value of the conditional upper bound is given by the set of equations in (42); for the bound in (46), the optimal tilting measure is given in (50).

IV. GENERALIZATION OF THE 1961 GALLAGER BOUND FOR PARALLEL CHANNELS AND ITS CONNECTION TO THE GENERALIZED DS2 BOUND

The 1961 Gallager bound for a single MBIOS channel was derived in [14], and a generalization of the bound for parallel MBIOS channels was proposed by Liu et al. [23]. In the following, we outline the derivation in [23] which serves as a preliminary step towards the discussion of its relation to the two versions of the generalized DS2 bound from Section III. In this section, we optimize the probability tilting measures which are related to the 1961 Gallager bound for J independent parallel channels in order to get the tightest bound within this form (hence, the optimization is carried w.r.t. J probability tilting measures). This optimization differs from the discussion in [23] where the authors choose some simple and sub-optimal tilting measures. By doing so, the authors in [23] derive bounds which are easier for numerical calculation, but the tightness of these bounds is loosened as compared to the improved bound which relies on the calculation of the J optimized tilting measures (this will be exemplified in Section VII for turbo-like ensembles).

A. Presentation of the Bound [23]

Consider a binary linear block code $\mathcal{C}$. Let $\underline{x}^m$ be the transmitted codeword and define the tilted ML metric

$$D_m(\underline{x}^{m'}, \underline{y}) \triangleq \ln \left(\frac{f_n^{(m)}(\underline{y})}{p_n(\underline{y}|\underline{x}^{m'})} \right) \tag{59}$$

where $f_n^{(m)}(\underline{y})$ is an arbitrary function which is positive if there exists $m' \neq m$ such that $p_n(\underline{y}|\underline{x}^{m'})$ is positive. If the code is ML decoded, an error occurs if for some $m' \neq m$

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

As noted in [36], $D_m(\cdot, \cdot)$ is in general not computable at the receiver. It is used here as a conceptual tool to evaluate the upper bound on the ML decoding error probability. The received set $\mathcal{Y}^n$ is expressed as a union of two disjoint subsets

$$
\begin{aligned}
\mathcal{Y}^n &= \mathcal{Y}_{\mathrm{g}}^n \cup \mathcal{Y}_{\mathrm{b}}^n \\
\mathcal{Y}_{\mathrm{g}}^n &\triangleq \left\{ \underline{y} \in \mathcal{Y}^n : D_m(\underline{x}^m, \underline{y}) \leq nd \right\} \\
\mathcal{Y}_{\mathrm{b}}^n &\triangleq \left\{ \underline{y} \in \mathcal{Y}^n : D_m(\underline{x}^m, \underline{y}) > nd \right\}
\end{aligned}
$$

where d is an arbitrary real number. The conditional ML decoding error probability can be expressed as the sum of two terms

$$P_{\mathrm{e}|m} = \mathrm{Prob}(\mathrm{error}, \underline{y} \in \mathcal{Y}_{\mathrm{b}}^n) + \mathrm{Prob}(\mathrm{error}, \underline{y} \in \mathcal{Y}_{\mathrm{g}}^n)$$

which is upper bounded by

$$P_{\mathrm{e}|m} \leq \mathrm{Prob}(\underline{y} \in \mathcal{Y}_{\mathrm{b}}^n) + \mathrm{Prob}(\mathrm{error}, \underline{y} \in \mathcal{Y}_{\mathrm{g}}^n) . \tag{60}$$

We use separate bounding techniques for the two terms in (60). Applying the Chernoff bound on the first term gives

$$P_1 \triangleq \mathrm{Prob}(\underline{y} \in \mathcal{Y}_{\mathrm{b}}^n) \leq \mathbf{E}\left(e^{sW}\right), \quad s \geq 0 \tag{61}$$

where

$$W \triangleq \ln\left(\frac{f_n^{(m)}(\underline{y})}{p_n(\underline{y}|\underline{x}^m)}\right) - nd . \tag{62}$$

Using a combination of the union and Chernoff bounds for the second term in the RHS of (60) gives

$$
\begin{aligned}
P_2 &\triangleq \mathrm{Prob}(\mathrm{error}, \underline{y} \in \mathcal{Y}_{\mathrm{g}}^n) \\
&= \mathrm{Prob}\left(D_m(\underline{x}^{m'}, \underline{y}) \leq D_m(\underline{x}^m, \underline{y}) \text{ for some } m' \neq m, \ \underline{y} \in \mathcal{Y}_{\mathrm{g}}^n \right) \\
&\leq \sum_{m' \neq m} \mathrm{Prob}\left(D_m(\underline{x}^{m'}, \underline{y}) \leq D_m(\underline{x}^m, \underline{y}), \ D_m(\underline{x}^m, \underline{y}) \leq nd \right) \\
&\leq \sum_{m' \neq m} \mathbf{E}\left(\exp(tU_{m'} + rW) \right), \quad t, r \leq 0
\end{aligned}
\tag{63}
$$

where, based on (59),

$$U_{m'} = D_m(\underline{x}^{m'}, \underline{y}) - D_m(\underline{x}^m, \underline{y}) = \ln\left(\frac{p_n(\underline{y}|\underline{x}^m)}{p_n(\underline{y}|\underline{x}^{m'})}\right) . \tag{64}$$

Consider a codeword of a binary linear block code $\mathcal{C}$ which is transmitted over J parallel MBIOS channels. Since the conditional error probability under ML decoding does not depend on the transmitted codeword, one can assume without loss of generality that the all-zero codeword is transmitted. As in Section III-A, we impose on the function $f_n^{(m)}(\underline{y})$ the restriction that it can be expressed in the product form

$$f_n^{(m)}(\underline{y}) = \prod_{j=1}^{J} \prod_{i \in \mathcal{I}(J)} f(y_i; j) . \tag{65}$$

For the continuation of the derivation, it is assumed that the functions $f(\cdot; j)$ are even, i.e., $f(y; j) = f(-y; j)$ for all $y \in \mathcal{Y}$. Plugging (25), (62), (64) and (65) into (61) and (63) we get

$$
\begin{aligned}
P_1 &\leq \sum_{\underline{y}} \left\{ \prod_{j=1}^{J} \prod_{i \in \mathcal{I}(j)} \left(\frac{f(y_i; j)}{p(y_i | 0; j)} \right)^s p(y_i | 0; j) \right\} e^{-nsd} \\
&= \prod_{j=1}^{J} \left\{ \left(\sum_{y \in \mathcal{Y}} p(y | 0; j)^{1-s} f(y; j)^s \right)^{n_j} \right\} e^{-nsd} \quad s \geq 0
\end{aligned}
\tag{66}
$$

$$
\begin{aligned}
P_2 &\leq \sum_{\underline{y}} \sum_{m' \neq m} \prod_{j=1}^{J} \prod_{i \in \mathcal{I}(j)} \left\{ \left(\frac{f(y_i; j)}{p(y_i | 0; j)} \right)^r p(y_i | 0; j) \left(\frac{p(y_i | 0; j)}{p(y_i | x_i^{(m')}; j)} \right)^t \right\} e^{-nrd} \\
&= \sum_{h_1=0}^{n_1} \cdots \sum_{h_J=0}^{n_J} \left\{ A_{h_1, \ldots, h_J} \prod_{j=1}^{J} \left[\sum_{y \in \mathcal{Y}} p(y | 0; j)^{1-r} f(y; j)^r \left(\frac{p(y | 0; j)}{p(y | 1; j)} \right)^t \right]^{h_j} \right. \\
&\qquad \left. \prod_{j=1}^{J} \left[\sum_{y \in \mathcal{Y}} p(y | 0; j)^{1-r} f(y; j)^r \right]^{n_j - h_j} \right\} e^{-nrd}, \quad t, r \leq 0
\end{aligned}
\tag{67}
$$

where as before, we use the notation $n_j \triangleq |\mathcal{I}(j)|$. Optimizing the parameter t gives the value in [14, Eq. (3.27)]

$$
t = \frac{r-1}{2}.
\tag{68}
$$

Let us define

$$
G(r; j) \triangleq \sum_{y} p(y | 0; j)^{1-r} f(y; j)^r
\tag{69}
$$

$$
Z(r; j) \triangleq \sum_{y} [p(y | 0; j) p(y | 1; j)]^{\frac{1-r}{2}} f(y; j)^r.
\tag{70}
$$

Substituting (68) into (67), combining the bounds on P_1 and P_2 in (66) and (67), and finally averaging over all possible channel assignments, we obtain

$$
\begin{aligned}
P_{\mathrm{e}} &\leq \mathbf{E} \left[\sum_{h=1}^{n} \sum_{\substack{0 \leq h_j \leq n_j \\ \sum h_j = h}} A_{h_1, \ldots, h_J} \prod_{j=1}^{J} [Z(r; j)]^{h_j} [G(r; j)]^{n_j - h_j} e^{-nrd} \right. \\
&\qquad \left. + \prod_{j=1}^{J} [G(s; j)]^{n_j} e^{-nsd} \right] \\
&= \sum_{\substack{n_j \geq 0 \\ \sum n_j = n}} \left\{ \sum_{h=1}^{n} \sum_{\substack{0 \leq h_j \leq n_j \\ \sum h_j = h}} A_{h_1, \ldots, h_J} \prod_{j=1}^{J} [Z(r; j)]^{h_j} [G(r; j)]^{n_j - h_j} e^{-nrd} \right. \\
&\qquad \left. + \prod_{j=1}^{J} [G(s; j)]^{n_j} e^{-nsd} \right\} P_{\underline{N}}(\underline{n}) \quad, \quad
\begin{array}{c}
r \leq 0 \\
s \geq 0 \\
-\infty < d < \infty
\end{array}.
\end{aligned}
\tag{71}
$$

Following the same procedure of averaging the bound over all possible assignments as in (30) and (31), we obtain

$$P_{\text{e}} \leq \sum_{h=1}^{n} \left\{ A_h \left(\sum_{j=1}^{J} \alpha_j Z(r;j) \right)^h \left(\sum_{j=1}^{J} \alpha_j G(r;j) \right)^{n-h} \right\} e^{-nrd}$$

$$+ \left(\sum_{j=1}^{J} \alpha_j G(s;j) \right)^n e^{-nsd} . \tag{72}$$

Finally, we optimize the bound in (72) over the parameter d which gives

$$P_{\text{e}} \leq 2^{H(\rho)} \left\{ \sum_{h=1}^{n} A_h \left[\sum_{j=1}^{J} \alpha_j Z(r;j) \right]^h \left[\sum_{j=1}^{J} \alpha_j G(r;j) \right]^{n-h} \right\}^\rho \left\{ \sum_{j=1}^{J} \alpha_j G(s;j) \right\}^{n(1-\rho)} \tag{73}$$

where $r \leq 0$, $s \geq 0$, and

$$\rho \triangleq \frac{s}{s-r} \quad , \quad 0 \leq \rho \leq 1 . \tag{74}$$

The bound in (73), originally derived in [23], is a natural generalization of the 1961 Gallager bound for parallel channels.

B. Connection to the Generalizations of the DS2 Bound

In this section we revisit the relations that exist between the DS2 bound and the 1961 Gallager bound, this time for the case of independent parallel channels. We will compare the 1961 Gallager bound with both versions of the DS2 bound presented in Sec. III. For the case of a single MBIOS channel, it was shown [10], [33], [36] that the DS2 bound is tighter than the 1961 Gallager bound.

This result easily extends to parallel channels, for the case of the second version of the DS2 bound which was derived in Sec. III-B using the perfect CSIR channel model. Under this model, the parallel-channel is expressed as a single MBIOS with output defined as the pair $b = (y, j)$. The results in [10], [33], [36] therefore apply directly to the CSIR model and can be used to show that the DS2 bound in (46) is tighter than the 1961 Gallager bound (73).

In this respect, the DS2 bound from Section III-A exhibits a slightly different behavior. In the remainder of this section, we provide analysis linking this bound with the 1961 Gallager bound. In what follows, we will see how a variation in the derivation of the Gallager bound leads to a form of the DS2 bound from Section III-A, up to a factor which varies between 1 and 2. To this end, we start from the point in the last section where the combination of the bounds in (66) and (67) is obtained. Rather than continuing as in the last section, we first optimize over the parameter d in the sum of the bounds on P_1 and P_2 in (66) and (67), yielding that

$$P_{\text{e}} \leq 2^{H(\rho)} \left\{ \sum_{h=1}^{n} \sum_{\substack{h_1,\ldots,h_j \\ \sum_j h_j = h}} A_{h_1,\ldots,h_j} \prod_{j=1}^{J} V(r,t;j)^{h_j} G(r;j)^{n_j-h_j} \right\}^\rho \prod_{j=1}^{J} G(s;j)^{n_j(1-\rho)}$$

$$= 2^{H(\rho)} \left\{ \sum_{h=1}^{n} \sum_{\substack{h_1,\ldots,h_j \\ \sum_j h_j = h}} A_{h_1,\ldots,h_j} \prod_{j=1}^{J} \left[V(r,t;j) G(s;j)^{\frac{1-\rho}{\rho}} \right]^{h_j} \right.$$

$$\left. \prod_{j=1}^{J} \left[G(r;j) G(s;j)^{\frac{1-\rho}{\rho}} \right]^{n_j-h_j} \right\}^\rho , \quad t, r \leq 0, \ s \geq 0$$

where

$$V(r,t;j) \triangleq \sum_{y} p(y|0;j)^{1-r} f(y;j)^r \left(\frac{p(y|0;j)}{p(y|1;j)} \right)^t \tag{75}$$

$G(\cdot; j)$ is introduced in (69) for $j = 1, \ldots, J$, and ρ is given in (74). Averaging the bound with respect to all possible channel assignments, we get for $0 \le \rho \le 1$

$$
\begin{aligned}
P_{\mathrm{e}} \le\ & 2^{H(\rho)} \sum_{\substack{n_j \ge 0 \\ \sum_j n_j = n}} \left\{ \left[\sum_{h=1}^{n} \sum_{\substack{h_1,\ldots,h_j \\ \sum_j h_j = h}} A_{h_1,\ldots,h_j} \prod_{j=1}^{J} \left[V(r,t;j) G(s;j)^{\frac{1-\rho}{\rho}} \right]^{h_j} \right. \right. \\
& \left. \left. \prod_{j=1}^{J} \left[G(r;j) G(s;j)^{\frac{1-\rho}{\rho}} \right]^{n_j - h_j} \right]^{\rho} P_{\underline{N}}(\underline{n}) \right\} \\
\le\ & 2^{H(\rho)} \left[\sum_{\substack{n_j \ge 0 \\ \sum_j n_j = n}} \sum_{h=1}^{n} \sum_{\substack{h_1,\ldots,h_j \\ \sum_j h_j = h}} A_{h_1,\ldots,h_j} P_{\underline{N}}(\underline{n}) \prod_{j=1}^{J} \left[V(r,t;j) G(s;j)^{\frac{1-\rho}{\rho}} \right]^{h_j} \right. \\
& \left. \prod_{j=1}^{J} \left[G(r;j) G(s;j)^{\frac{1-\rho}{\rho}} \right]^{n_j - h_j} \right]^{\rho} \quad (76)
\end{aligned}
$$

where we invoked Jensen's inequality in the last step. Following the same steps as in (28)–(31), we get

$$
\begin{aligned}
P_{\mathrm{e}} \le\ & 2^{H(\rho)} \left[\sum_{h=1}^{n} A_h \left(\sum_{j=1}^{J} \alpha_j V(r,t;j) G(s;j)^{\frac{1-\rho}{\rho}} \right)^{h} \right. \\
& \left. \left(\sum_{j=1}^{J} \alpha_j G(r;j) G(s;j)^{\frac{1-\rho}{\rho}} \right)^{n-h} \right]^{\rho}, \quad (77)
\end{aligned}
$$

where from (68), (69), (74) and (75)

$$
\begin{aligned}
G(s;j) &= \sum_{y} p(y|0;j) \left(\frac{f(y;j)}{p(y|0;j)} \right)^{s} \\
G(r;j) &= \sum_{y} p(y|0;j) \left(\frac{f(y;j)}{p(y|0;j)} \right)^{s(1-\frac{1}{\rho})} \\
V(r,t;j) &= \sum_{y} p(y|0;j) \left(\frac{f(y;j)}{p(y|0;j)} \right)^{s(1-\frac{1}{\rho})} \left(\frac{p(y|0;j)}{p(y|1;j)} \right)^{t}. \quad (78)
\end{aligned}
$$

Setting $\lambda = -t$, and substituting in (78) the following relation between the Gallager tilting measures and the un-normalized tilting measures in the DS2 bound

$$
g(y;j) \triangleq \left(\frac{f(y;j)}{p(y|0;j)} \right)^{s}, \quad j = 1, 2, \ldots, J \quad (79)
$$

we obtain

$$
\begin{aligned}
P_{\mathrm{e}} \le\ & 2^{H(\rho)} \left\{ \sum_{h=0}^{n} A_h \left[\sum_{j=1}^{J} \alpha_j \left(\sum_{y} g(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{1-\lambda} p(y|1;j)^{\lambda} \right) \right. \right. \\
& \left. \left(\sum_{y} g(y;j) p(y|0;j) \right)^{\frac{1-\rho}{\rho}} \right]^{h} \left[\sum_{j=1}^{J} \alpha_j \left(\sum_{y} g(y;j)^{1-\frac{1}{\rho}} p(y|0;j) \right) \right. \\
& \left. \left. \left(\sum_{y} g(y;j) p(y|0;j) \right)^{\frac{1-\rho}{\rho}} \right]^{n-h} \right\}^{\rho}, \quad 0 \le \rho \le 1 \quad (80)
\end{aligned}
$$

which coincides with the form of the DS2 bound given in (33) (up to the factor $2^{H(\rho)}$ which lies between 1 and 2), for those un-normalized tilting measures $g(\cdot; j)$ such that the resulting functions $f(\cdot; j)$ in (79) are *even*.

Discussion. The derivation of the 1961 Gallager bound involves first the averaging of the bound in (71) over all possible channel assignments and then the optimization over the parameter d in (72). To show a connection to the DS2 bound in (33), we had first optimized over d and then obtained the bound averaged over all possible channel assignments. The difference between the two approaches is that in the latter, Jensen's inequality had to be used in (76) to continue the derivation (because the expectation over all possible channel assignments was performed on an expression raised to the ρ-th power) which resulted in the DS2 bound, whereas in the derivation of [23], the need for Jensen's inequality was circumvented due to the linearity of the expression in (71). We note that Jensen's inequality was also used for the direct derivation of the DS2 bound in (31); this use of Jensen's inequality hinders the tightness of this bound to the point where we cannot determine if it is tighter than the 1961 Gallager bound or not. For the special case of $J = 1$, both versions of the DS2 bound degenerate to the standard DS2 bound from Sec. II-D. In this case, as in the case of the DS2 bound from Section III-B, the DS2 bound is tighter than the 1961 Gallager bound (as noted in [36]) due to the following reasons:

- For the 1961 Gallager bound, it is required that $f(\cdot; j)$ be even. This requirement inhibits the optimization of $\psi(\cdot; j)$ in Section III because the optimal choice of $\psi(\cdot; j)$ given in (42) leads to functions $f(\cdot; j)$ which are not even. The exact form of $f(\cdot; j)$ which stems from the optimal choice of $\psi(\cdot; j)$ is detailed in Appendix A.1.
- The absence of the factor $2^{H(\rho)}$ (which is greater than 1) in both versions of the DS2 bound implies their superiority. Naturally, this factor is of minor importance since we are primarily interested in the exponential tightness of these bounds.

It should be noted that, as in the case of $J = 1$, the optimization over the DS2 tilting measure is still over a larger set of functions as compared to the 1961 Gallager tilting measure; hence, the derivation appearing in this section of the DS2 bound in (33) from the 1961 Gallager bound *only gives an expression of the same form and not the same upper bound* (disregarding the $2^{H(\rho)}$ constant).

C. Optimized Tilting Measures for the Generalized 1961 Gallager Bound

We derive in this section optimized tilting measures for the 1961 Gallager bound. These optimized tilting measures are derived for random coding, and for the case of constant Hamming weight codes. The 1961 Gallager bound will be used later in conjunction with these optimized tilting measures in order to get an upper bound on the decoding error probability of an arbitrary binary linear block code. To this end, such a code is partitioned to constant Hamming weight subcodes (where each subcode also includes the all-zero codeword), and a union bound is used in conjunction with the calculation of the conditional error probability of each subcode, given that the all-zero codeword is transmitted. Using these optimized tilting measures improves the tightness of the resulting bound, as exemplified in the continuation of this paper.

1) Tilting Measures for Random Codes: Consider the ensemble of fully random binary block codes of length n. Substituting the appropriate weight enumerator (given in (15)) into (72), we get

$$
\begin{aligned}
P_{\mathrm{e}} \leq \; & 2^{-n(1-R)} \left\{ \frac{1}{2} \sum_{j=1}^{J} \alpha_j \sum_y \left[p(y|0;j)^{\frac{1-r}{2}} + p(y|1;j)^{\frac{1-r}{2}} \right]^2 f(y;j)^r \right\}^n e^{-nrd} \\
& + \left\{ \frac{1}{2} \sum_{j=1}^{J} \alpha_j \sum_y \left(p(y|0;j)^{1-s} + p(y|1;j)^{1-s} \right) f(y;j)^s \right\}^n e^{-nsd} , \quad \begin{matrix} r \leq 0 \\ s \geq 0 \\ d \in \mathbb{R} \end{matrix}
\end{aligned}
\tag{81}
$$

where we rely on (69) and (70), use the symmetry of the channels and the fact that we require the functions $f(\cdot; j)$ $(j = 1, \ldots, J)$ to be even. To optimize (81) over all possible tilting measures, we apply calculus of variations. This procedure gives the following equation:

$$
\begin{aligned}
& \sum_{j=1}^{J} \alpha_j \left(p(y|0;j)^{\frac{1-r}{2}} + p(y|1;j)^{\frac{1-r}{2}} \right)^2 f(y;j)^{r-1} \\
& - L \sum_{j=1}^{J} \alpha_j \left(p(y|0;j)^{1-s} + p(y|1;j)^{1-s} \right)^2 f(y;j)^{s-1} = 0 \quad \forall y.
\end{aligned}
$$

where $L \in \mathbb{R}$ is a Lagrange multiplier. This equation is satisfied for tilting measures which are given in the form

$$f(y;j) = K \left\{ \frac{\left(p(y|0;j)^{\frac{1-r}{2}} + p(y|1;j)^{\frac{1-r}{2}}\right)^2}{p(y|0;j)^{1-s} + p(y|1;j)^{1-s}} \right\}^{\frac{1}{s-r}} \qquad K \in \mathbb{R}. \tag{82}$$

This forms a natural generalization of the tilting measure given in [14, Eq. (3.41)] for a single MBIOS channel. We note that the scaling factor K may be omitted as it cancels out when we substitute (82) in (73).

2) Tilting Measures for Constant Hamming Weight Codes: The distance spectrum of a constant Hamming weight code is given by

$$A_{h'} = \begin{cases} 1, & \text{if } h' = 0 \\ A_h, & \text{if } h' = h \\ 0, & \text{otherwise} \end{cases} \tag{83}$$

Substituting this into (73) and using the symmetry of the component channels and the fact that the tilting measures $f(\cdot;j)$ are required to be even, we get

$$\begin{aligned}
P_{e|0}(h) &\leq 2^{H(\rho)} A_h^{\rho} \left\{ \sum_{j=1}^{J} \alpha_j \sum_y [p(y|0;j)p(y|1;j)]^{\frac{1-r}{2}} f(y;j)^r \right\}^{h\rho} \\
&\cdot \left\{ \sum_{j=1}^{J} \frac{\alpha_j}{2} \sum_y \left[p(y|0;j)^{1-r} + p(y|1;j)^{1-r} \right] f(y;j)^r \right\}^{(n-h)\rho} \\
&\cdot \left\{ \sum_{j=1}^{J} \frac{\alpha_j}{2} \sum_y \left[p(y|0;j)^{1-s} + p(y|1;j)^{1-s} \right] f(y;j)^s \right\}^{n(1-\rho)}, \\
&\qquad\qquad\qquad\qquad r \leq 0, \ s \geq 0, \ \rho = \frac{s}{s-r}.
\end{aligned} \tag{84}$$

Applying calculus of variations to (84) yields (see Appendix A.2 for some additional details) that the following condition should be satisfied for all values of $y \in \mathcal{Y}$:

$$\sum_{j=1}^{J} \alpha_j \left\{ \left[p(y|0;j)^{1-s} + p(y|1;j)^{1-s} \right] f(y;j)^{s-r} + K_1 \left[p(y|0;j)p(y|1;j) \right]^{\frac{1-r}{2}} \right. \tag{85}$$

$$\left. + K_2 \left[p(y|0;j)^{1-r} + p(y|1;j)^{1-r} \right] \right\} = 0$$

where $K_1, K_2 \in \mathbb{R}$. This condition is satisfied if we require

$$\begin{aligned}
&\left[p(y|0;j)^{1-s} + p(y|1;j)^{1-s} \right] f(y;j)^{s-r} + K_1 \left[p(y|0;j)p(y|1;j) \right]^{\frac{1-r}{2}} \\
&+ K_2 \left[p(y|0;j)^{1-r} + p(y|1;j)^{1-r} \right] \equiv 0, \qquad \forall y \in \mathcal{Y}, \ j = 1, \ldots, J.
\end{aligned}$$

The optimized tilting measures can therefore be expressed in the form

$$\begin{aligned}
f(y;j) = &\left\{ \frac{c_1 \left(p(y|0;j)^{\frac{1-s(1-\rho^{-1})}{2}} + p(y|1;j)^{\frac{1-s(1-\rho^{-1})}{2}} \right)^2}{p(y|0;j)^{1-s} + p(y|1;j)^{1-s}} + \right. \\
&\left. \frac{d_1 \left(p(y|0;j)^{1-s(1-\rho^{-1})} + p(y|1;j)^{1-s(1-\rho^{-1})} \right)}{p(y|0;j)^{1-s} + p(y|1;j)^{1-s}} \right\}^{\frac{\rho}{s}}, \quad \begin{array}{c} c_1, d_1 \in \mathbb{R} \\ s \geq 0 \\ 0 \leq \rho \leq 1 \end{array}
\end{aligned} \tag{86}$$

where we have used (74). This form is identical to the optimal tilting measure for random codes if we set $d_1 = 0$. It is possible to scale the parameters c_1 and d_1 without affecting the 1961 Gallager bound (i.e., the ratio $\frac{c_1}{d_1}$ cancels out when we substitute (86) in (73)). Furthermore, we note that regardless of the values of c_1 and d_1, the resulting tilting measures are even functions, as required in the derivation of the 1961 Gallager bound.

For the simplicity of the optimization, we wish to reduce the infinite intervals in (86) to finite ones. It is shown in [31, Appendix A] that the optimization of the parameter s can be reduced to the interval $[0, 1]$ without loosening the tightness of the bound. Furthermore, the substitution $c \triangleq \frac{c_1 + 2d_1}{2c_1 + 3d_1}$, as suggested in [31, Appendix B], enables one to express the optimized tilting measure in (86) using an equivalent form where the new parameter c lies in the interval $[0, 1]$. The numerical optimization of the bound in (86) is therefore taken over the range of parameters $0 \leq \rho \leq 1$, $0 \leq s \leq 1$, $0 \leq c \leq 1$. Based on the calculations in [31, Appendices A, B], the functions $f(\cdot; j)$ get the equivalent form

$$
\begin{aligned}
f(y; j) &= \left\{ \frac{(1-c) \left(p(y|0; j)^{\frac{1-s(1-\rho^{-1})}{2}} - p(y|1; j)^{\frac{1-s(1-\rho^{-1})}{2}} \right)^2}{p(y|0; j)^{1-s} + p(y|1; j)^{1-s}} \right. \\
&\left. + \frac{2c \left(p(y|0; j) p(y|1; j) \right)^{\frac{1-s(1-\rho^{-1})}{2}}}{p(y|0; j)^{1-s} + p(y|1; j)^{1-s}} \right\}^{\frac{\rho}{s}}, \quad (\rho, s, c) \in [0, 1]^3.
\end{aligned}
\tag{87}
$$

By reducing the optimization of the three parameters over the unit cube, the complexity of the numerical process is reduced to an acceptable level.

D. Statement of the Main Result Derived in Section IV

The analysis in this section leads to the following theorem:

Theorem 2 (Generalized 1961 Gallager bound for parallel channels): Consider the transmission of binary linear block codes (or ensembles) over a set of J independent parallel MBIOS channels. Following the notation in Theorem 1, the generalization of the 1961 Gallager bound in (73) provides an upper bound on the ML decoding error probability when the bound is taken over the whole code (as originally derived in [23]). By partitioning the code into constant Hamming-weight subcodes, the generalized 1961 Gallager bound on the conditional ML decoding error probability of an arbitrary subcode (given that the all-zero codeword is transmitted) is provided by (84), and (34) forms an upper bound on the block error probability of the whole code (or ensemble). For an arbitrary constant Hamming weight subcode, the optimized set of non-negative and even functions $\{f(\cdot; j)\}_{j=1}^{J}$ which attains the minimal value of the conditional bound in (84), is given by (87); this set of functions is subject to a three-parameter optimization over a cube of unit length (see (87)).

V. Special Cases of the Generalized DS2 Bound for Independent Parallel Channels

In this section, we rely on the two versions of the generalized DS2 bound for independent parallel MBIOS channels, as presented in Sections III-A and III-B, and apply them in order to re-derive some of the bounds which were originally derived by Liu et al. [23]. The derivation in [23] is based on the 1961 Gallager bound from Section IV-A, and the authors choose particular and sub-optimal tilting measures in order to get closed form bounds (in contrast to the optimized tilting measures in Section IV-C which lead to more complicated bounds in terms of their numerical computation). In this section, we follow the same approach in order to re-derive some of their bounds as particular cases of the two generalized DS2 bounds (i.e., we choose some particular tilting measures rather than the optimized ones). In some cases, we re-derive the bounds from [23] as special cases of the generalized DS2 bound, or alternatively, obtain some modified bounds as compared to [23].

A. Union-Bhattacharyya Bound in Exponential Form

As in the case of a single channel, it is a special case of both versions of the DS2 and the 1961 Gallager bound. By substituting $r = 0$ in the Gallager bound or $\rho = 1, \lambda = 0.5$ in both versions of the DS2 bound, we get

$$
P_{\mathrm{e}} \leq \sum_{h=1}^{n(1-R)} A_h \gamma^h
\tag{88}
$$

where γ is given by (3) and denotes the average Bhattacharyya parameter of J independent parallel channels. Note that this bound is given in exponential form, i.e., as in the single channel case, it doesn't use the exact expression for the pairwise error probability between two codewords of Hamming distance h. For the case of the binary-input AWGN, a tighter version which uses the Q-function to express the exact pairwise error probability is presented in Appendix C.

B. The Sphere Bound for Parallel AWGN Channels

The simplified sphere bound is an upper bound on the ML decoding error probability for the binary-input AWGN channel. In [23], the authors have obtained a parallel-channel version of the sphere bound by making the substitution $f(y;j) = \frac{1}{\sqrt{2\pi}}$ in the 1961 Gallager bound. We will show that this version is also a special case of both versions of the parallel-channel DS2 bound. By using the relation (79), between Gallager's tilting measure and the un-normalized DS2 tilting measure, we get

$$g(y;j) = \left(\frac{f(y;j)}{p(y|0;j)}\right)^s = \exp\left(\frac{s(y+\sqrt{2\nu_j})^2}{2}\right)$$

so that

$$\int_{-\infty}^{+\infty} g(y;j)p(y|0;j)\,dy = \frac{1}{\sqrt{1-s}}$$

$$\int_{-\infty}^{+\infty} g(y;j)^{1-\frac{1}{\rho}}p(y|0;j)\,dy = \frac{1}{\sqrt{1-s\left(1-\frac{1}{\rho}\right)}}$$

$$\int_{-\infty}^{+\infty} g(y;j)^{1-\frac{1}{\rho}}p(y|0;j)^{1-\lambda}p(y|1;j)^\lambda\,dy = \frac{e^{\nu_j\left(1-s\left(1-\frac{1}{\rho}\right)\right)}}{\sqrt{1-s\left(1-\frac{1}{\rho}\right)}}.$$

By introducing the two new parameters $\beta = 1 - s\left(1 - \frac{1}{\rho}\right)$ and $\lambda = \frac{\beta}{2}$ we get

$$\int_{-\infty}^{+\infty} g(y;j)p(y|0;j)dy = \sqrt{\frac{1-\rho}{1-\beta\rho}}$$

$$\int_{-\infty}^{+\infty} g(y;j)^{1-\frac{1}{\rho}}p(y|0;j)dy = \beta^{-\frac{1}{2}} \tag{89}$$

$$\int_{-\infty}^{+\infty} g(y;j)^{1-\frac{1}{\rho}}p(y|0;j)^{1-\lambda}p(y|1;j)^\lambda dy = \frac{\gamma_j^\beta}{\sqrt{\beta}}, \quad \gamma_j \triangleq e^{-\nu_j}.$$

Next, by plugging (89) into the DS2 bound in (33), we get

$$P_{\rm e} \leq \left\{\sum_{h=0}^{n} A_h \left(\sum_{j=1}^{J}\alpha_j\gamma_j^\beta\right)^h \beta^{-\frac{n}{2}}\right\}^\rho \left(\frac{1-\rho}{1-\beta\rho}\right)^{\frac{n(1-\rho)}{2}}, \quad \begin{array}{c} 0 \leq \rho \leq 1 \\ 1 \leq \beta \leq \frac{1}{\rho} \end{array}. \tag{90}$$

The same expression may be obtained by plugging (89) into the DS2 bound in (47). This bound is identical to the parallel-channel simplified sphere bound in [23, Eq. (24)], except that it provides a slight improvement due to the absence of the factor $2^{H(\rho)}$ which appears in [23, Eq. (24)] (a factor bounded between 1 and 2).

C. Generalizations of the Shulman-Feder Bound for Parallel Channels

In this sub-section, we present two generalizations of the Shulman and Feder (SF) bound, where both bounds apply to independent parallel channels. The first bound was previously obtained by Liu et al. [23] as a special case of the generalization of the 1961 Gallager bound and will be shown to be a special case of the DS2 bound from Section III-B, and the second bound follows as a particular case of the DS2 bound from Section III-A for independent parallel channels.

By substituting in (73) the tilting measure and the parameters (see [23, Eq. (28)])

$$f(y;j) = \left(\frac{1}{2}p(y|0;j)^{\frac{1}{1+\rho}} + \frac{1}{2}p(y|1;j)^{\frac{1}{1+\rho}}\right)^{1+\rho}$$

$$r = -\frac{1-\rho}{1+\rho}, \quad s = \frac{\rho}{1+\rho}, \quad 0 \leq \rho \leq 1 \tag{91}$$

straightforward calculations for MBIOS channels give the following bound which was originally introduced in [23, Lemma 2]:

$$P_e \leq 2^{H(\rho)} 2^{nR\rho} \left(\max_{1 \leq h \leq n} \frac{A_h}{2^{-n(1-R)} \binom{n}{h}} \right)^{\rho} \left\{ \sum_{j=1}^{J} \alpha_j \left(\sum_y \frac{1}{2} p(y|0;j)^{\frac{1}{1+\rho}} + \frac{1}{2} p(y|1;j)^{\frac{1}{1+\rho}} \right)^{1+\rho} \right\}^n. \quad (92)$$

Due to the natural connection between the DS2 bound in Section III-B and the 1961 Gallager bound for parallel channels (see the discussion in Sec. IV-B), the generalized SF bound is also a special case of the former bound. The tilting measure which should be used in this case to show the connection has already appeared in (56) (as a part of the discussion Sec. III-C on the random coding version of this bound) and it reads

$$g(y;j) = \left[\frac{1}{2} p(y|0;j)^{\frac{1}{1+\rho}} + \frac{1}{2} p(y|1;j)^{\frac{1}{1+\rho}} \right]^{\rho} p(y|0;j)^{-\frac{\rho}{1+\rho}}.$$

The result is the same as the bound in (92) except for the absence of the factor $2^{H(\rho)}$.

Considering the generalization of the DS2 bound in Section III-A, it is possible to start from Eq. (33) and take the maximum distance spectrum term out of the sum. This gives the bound

$$\begin{aligned} P_e \quad \leq \quad & 2^{-n(1-R)\rho} \left(\max_{1 \leq h \leq n} \frac{A_h}{2^{-n(1-R)} \binom{n}{h}} \right)^{\rho} \left\{ \sum_{j=1}^{J} \alpha_j \left[\sum_y g(y;j) p(y|0;j) \right]^{\frac{1-\rho}{\rho}} \right. \\ & \left. \cdot \left[\sum_y p(y|0;j) g(y;j)^{1-\frac{1}{\rho}} \left(1 + \left(\frac{p(y|1;j)}{p(y|0;j)} \right)^{\lambda} \right) \right] \right\}^{n\rho}, \quad 0 \leq \rho \leq 1. \end{aligned} \quad (93)$$

Using the J un-normalized tilting measures from (56) and setting $\lambda = \frac{1}{1+\rho}$ in (93), gives the following bound due to the symmetry at the channel outputs:

$$\begin{aligned} P_e \quad \leq \quad & 2^{nR\rho} \left(\max_{1 \leq h \leq n} \frac{A_h}{2^{-n(1-R)} \binom{n}{h}} \right)^{\rho} \\ & \left\{ \sum_{j=1}^{J} \alpha_j \left[\left(\sum_y \frac{1}{2} p(y|0;j)^{\frac{1}{1+\rho}} + \frac{1}{2} p(y|1;j)^{\frac{1}{1+\rho}} \right)^{1+\rho} \right]^{\frac{1}{\rho}} \right\}^{n\rho}, \quad 0 \leq \rho \leq 1 \end{aligned} \quad (94)$$

which forms another possible generalization of the SF bound for independent parallel channels. Clearly, unless $J = 1$ (referring to the case of a single MBIOS channel), this bound is exponentially looser than the one in (92). The fact that the bound in (94) is exponentially looser than the bound in (92) follows from the use of Jensen's inequality for the derivation of the first version of the DS2 bound (see the move from (28) to (29)).

D. Modified Shulman-Feder Bound for Independent Parallel Channels

It is apparent from the form of the SF bound that its exponential tightness depends on the quantity

$$\max_{1 \leq h \leq n} \frac{A_h}{2^{-n(1-R)} \binom{n}{h}} \quad (95)$$

which measures the maximal ratio of the distance spectrum of the considered binary linear block code (or ensemble) and the average distance spectrum of fully random block codes with the same rate and block length. One can observe from Fig. 3 that this ratio may be quite large for a non-negligible portion of the normalized Hamming weights, thus undermining the tightness of the SF bound. The idea of the *Modified Shulman-Feder* (MSF) bound is to split the set of non-zero normalized Hamming weights $\Psi_n \triangleq \left\{ \frac{1}{n}, \frac{2}{n}, \ldots, 1 \right\}$ into two disjoint subsets Ψ_n^+ and Ψ_n^- where the union bound is used for the codewords with normalized Hamming weights within the set Ψ_n^+, and the SF bound is used for the remaining codewords. This concept was originally applied to the ML analysis of ensembles of LDPC codes by Miller and Burshtein [26]. Typically, the set Ψ_n^+ consists of low and high Hamming weights, where the ratio in (95) between the distance spectra and the binomial distribution appears to be quite large for typical code ensembles of linear codes; the set Ψ_n^- is the complementary set which includes medium values of the normalized

Hamming weight. The MSF bound for a given partitioning Ψ_n^-, Ψ_n^+ is introduced in [23, Lemma 3], and gets the form

$$
P_{\mathrm{e}} \leq \sum_{h: \frac{h}{n} \in \Psi_n^+} A_h \gamma^h + 2^{H(\rho)} 2^{nR\rho} \left(\max_{h: \frac{h}{n} \in \Psi_n^-} \frac{A_h}{2^{-n(1-R)} \binom{n}{h}} \right)^\rho
$$
$$
\cdot \left\{ \sum_{j=1}^J \alpha_j \left(\sum_y \frac{1}{2} p(y|0;j)^{\frac{1}{1+\rho}} + \frac{1}{2} p(y|1;j)^{\frac{1}{1+\rho}} \right)^{1+\rho} \right\}^n
\tag{96}
$$

where γ is introduced in (3), and $0 \leq \rho \leq 1$. Liu et al. prove that in the limit where the block length tends to infinity, the optimal partitioning of the set of non-zero normalized Hamming weights to two disjoint subsets Ψ_n^- and Ψ_n^+ is given by (see [23, Eq. (42)])

$$
\delta \in \left\{ \begin{array}{ll} \Psi_n^+ & \text{if} \quad -\delta \ln \gamma \geq H(\delta) + (\overline{I} - 1) \ln 2 \\ \Psi_n^- & \text{otherwise} \end{array} \right.
\tag{97}
$$

where

$$
\overline{I} \triangleq \sum_{j=1}^J \frac{\alpha_j}{2} \sum_{x \in \{-1,1\}} \sum_y p(y|x;j) \log_2 \frac{p(y|x;j)}{1/2 \sum_{x' \in \{-1,1\}} p(y|x';j)}
$$

designates the average mutual information under the assumption of equiprobable binary inputs. Note that for finite block lengths, even with the same partitioning as above, the first term in the RHS of (96) can be tightened by replacing the Bhattacharyya bound with the exact expression for the average pairwise error probability between two codewords of Hamming distance h. Referring to parallel binary-input AWGN channels, the exact pairwise error probability is given in (C.5), thus providing the following tightened upper bound:

$$
P_{\mathrm{e}} \leq \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \left\{ \sum_{h: \frac{h}{n} \in \Psi_n^+} A_h \left[\sum_{j=1}^J \alpha_j e^{-\frac{\nu_j}{\sin^2 \theta}} \right]^h \mathrm{d}\theta \right\}
$$
$$
+ 2^{H(\rho)} 2^{nR\rho} \left(\max_{h: \frac{h}{n} \in \Psi_n^-} \frac{A_h}{2^{-n(1-R)} \binom{n}{h}} \right)^\rho
$$
$$
\cdot \left\{ \sum_{j=1}^J \alpha_j \left(\sum_y \frac{1}{2} p(y|0;j)^{\frac{1}{1+\rho}} + \frac{1}{2} p(y|1;j)^{\frac{1}{1+\rho}} \right)^{1+\rho} \right\}^n.
\tag{98}
$$

On the selection of a suitable partitioning of the set Ψ_n in (98): The asymptotic partitioning suggested in (97) typically yields that the union bound is used for low and high values of normalized Hamming weights; for these values, the distance spectrum of ensembles of turbo-like codes deviates considerably from the binomial distribution (referring to the ensemble of fully random block codes of the same block length and rate). Let δ_l and δ_r be the smallest and largest normalized Hamming weights, respectively, referring to the range of values of δ in (97) so that $\Psi_n^- \triangleq \left\{ \delta_l, \delta_l + \frac{1}{n}, \ldots, \delta_r \right\}$, and $\Psi_n^+ \triangleq \left\{ \frac{1}{n}, \frac{2}{n}, \ldots, \delta_l - \frac{1}{n} \right\} \cup \left\{ \delta_r + \frac{1}{n}, \delta_r + \frac{2}{n}, \ldots, 1 \right\}$ are the sets of normalized Hamming weights. The subsets Ψ_n^+ and Ψ_n^- refer to the discrete values of normalized Hamming weights for which the union bound in its exponential form is superior to the SF bound and vice versa, respectively (see (96)). Our numerical experiments show that for finite-length codes (especially, for codes of small and moderate block lengths), this choice of δ_l and δ_r often happens to be sub-optimal in the sense of minimizing the overall upper bounds in (96) and (98). This happens because for $\delta = \delta_l$ (which is the left endpoint of the interval for which the SF bound is calculated), the ratio of the average distance spectrum of the considered ensemble and the one which corresponds to fully random block codes is rather large, so the second term in the right-hand-sides of (96) and (98) corresponding to the contribution of the SF bound to the overall bound is considerably larger than the first term which refers to the union bound. Therefore, for finite-length codes, the following algorithm is proposed to optimize the partition $\Psi_n = \Psi_n^+ \cup \Psi_n^-$:

1) Select initial values δ_{l_0} and δ_{r_0} (for δ_l and δ_r) via (97). If there are less than two solutions to the equation $-\delta \ln \gamma = H(\delta) + (\overline{I} - 1) \ln 2$, select $\Psi_n^+ = \Psi_n$, $\Psi_n^- = \phi$ as the empty set.

2) Optimize the value of δ_l by performing a linear search in the range $[\delta_{l_0}, \delta_{r_0}]$ and finding the value of δ_l which minimizes the overall bound in the RHS of (98).

The optimal partition for the MSF bound yields the tightest bound within this family. This bound yields the *largest* MSF attainable channel region, and is therefore termed the LMSF bound. The algorithm above is applied to the calculation of the LMSF bound for finite-length codes (see, e.g., 4(b) in p. 32).

Clearly, an alternative and slightly tighter version of the MSF bound can be obtained from the DS2 bound from Section III-B for parallel channels where the difference will be in the absence of the $2^{H(\rho)}$ constant. We address the LMSF bound in Section VII, where for various ensembles of turbo-like codes, its tightness is compared with that of both versions of generalized DS2 and Gallager bounds.

VI. INNER BOUNDS ON ATTAINABLE CHANNEL REGIONS FOR ENSEMBLES OF GOOD BINARY LINEAR CODES TRANSMITTED OVER PARALLEL CHANNELS

In this section, we consider inner bounds on the attainable channel regions for ensembles of good binary linear codes (e.g., turbo-like codes) whose transmission takes place over independent parallel channels. The computation of these regions follows from the upper bounds on the ML decoding error probability we have obtained in Sections III and IV (see Theorems 1 and 2), referring here to the asymptotic case where we let the block length tend to infinity.

Let us consider an ensemble of binary linear codes, and assume that the codewords of each code are transmitted with equal probability. A J-tuple of transition probabilities characterizing a parallel channel is said to be an *attainable channel point* with respect to a code ensemble $\mathcal{C}$ if the average ML decoding error probability vanishes as we let the block length tend to infinity. The *attainable channel region* of an ensemble whose transmission takes place over parallel channels is defined as the closure of the set of attainable channel points. We will focus here on the case where each of the J independent parallel channels can be described by a single real parameter, i.e., the attainable channel region is a subset of $\mathbb{R}^J$; the boundary of the attainable region is called the *noise boundary* of the channel. Since the exact decoding error probability under ML decoding is in general unknown, then similarly to [23], we evaluate inner bounds on the attainable channel regions whose calculation is based on upper bounds on the ML decoding error probability.

In [23, Section 4], Liu et al. have used special cases of the 1961 Gallager bound to derive a simplified algorithm for calculating inner bounds on attainable channel regions. As compared to the bounds introduced in [23], the improvement in the tightness of the bounds presented in Theorems 1 and 2 is expected to enlarge the corresponding inner bounds on the attainable channel regions. Our numerical results referring to inner bounds on attainable channel regions are based on the following theorem:

Theorem 3 (Inner bounds on the attainable channel regions for parallel channels): Let us assume that the transmission of a sequence of binary linear block codes (or ensembles) $\{[\mathcal{C}(n)]\}$ takes place over a set of J parallel MBIOS channels. Assume that the bits are randomly assigned to these channels, so that every bit is transmitted over a single channel and the a-priori probability for transmitting a bit over the j-th channel is α_j (where $\sum_{j=1}^{J} \alpha_j = 1$ and $\alpha_j \geq 0$ for $j \in \{1, \ldots, J\}$). Let $\{A_h^{[\mathcal{C}(n)]}\}$ designate the (average) distance spectrum of the sequence of codes (or ensembles), $r^{[\mathcal{C}]}(\delta)$ designate the asymptotic exponent of the (average) distance spectrum, and

$$\gamma_j \triangleq \sum_{y \in \mathcal{Y}} \sqrt{p(y|0;j)p(y|1;j)}, \quad j \in \{1, \ldots, J\}$$

designate the Bhattachryya constants of the channels. Assume that the following conditions hold:

1)
$$\inf_{\delta_0 < \delta \leq 1} E^{\mathrm{DS2}_1}(\delta) > 0, \quad \forall \, \delta_0 \in (0,1) \tag{99}$$

where, for $0 < \delta \leq 1$, $E^{\mathrm{DS2}_1}(\delta)$ is calculated from (37) by maximizing w.r.t. λ, ρ ($\lambda \geq 0$ and $0 \leq \rho \leq 1$) and the probability tilting measures $\{\psi(\cdot;j)\}_{j=1}^{J}$.

2) The inequality
$$\limsup_{\delta \to 0} \frac{r^{[\mathcal{C}]}(\delta)}{\delta} < -\ln\left(\sum_{j=1}^{J} \alpha_j \gamma_j\right) \tag{100}$$

is satisfied, where the sum inside the logarithm designates the average Bhattacharrya constant over the J parallel channels, and $r^{[\mathcal{C}]}(\delta)$ designates the asymptotic growth rate of the distance spectrum as defined in (14).

3) There exists a sequence $\{D_n\}$ of natural numbers tending to infinity with increasing n so that

$$\limsup_{n \to \infty} \sum_{h=1}^{D_n} A_h^{[\mathcal{C}(n)]} = 0 \tag{101}$$

4) The exponent of the distance spectrum $r^{[\mathcal{C}(n)]}(\delta)$ converges uniformly in $\delta \in [0, 1]$ to its asymptotic limit.

Then, the J-tuple vector of parameters characterizing these channels lies within the attainable channel region under ML decoding.

Proof: The reader is referred to Appendix B. ∎

Discussion: We note that conditions 3 and 4 in Theorem 3 are similar to the last two conditions in [22, Theorem 2.3]. Condition 2 above happens to be a natural generalization of the second condition in [22, Theorem 2.3], thus generalizing the single channel case to a set of parallel channels. The distinction between [22, Theorem 2.3] which relates to typical-pairs decoding over a single channel and the statement in Theorem 3 for ML decoding over a set of independent parallel channels lies mainly in the first condition of both theorems.

A similar result which involves the generalized 1961 Gallager bound for parallel channels can be proven in the same way by replacing the first condition with an equivalent relation involving the exponent of the 1961 Gallager bound maximized over its parameters, instead of the error exponent of the DS2 bound.

The difference of our results from those presented in [23] stems from the fact that we rely here on the generalized DS2 bound and the 1961 Gallager bound with their related *optimized tilting measures*, and not on particular cases of the latter bound. These optimizations which are carried over the tilting measures of both bounds provide tighter bounds as compared to the bounds introduced in [23, Sections 4 and 5] which follow from the particular choices of the tilting measures for the generalized 1961 Gallager bound.

We later exemplify our inner bounds on the attainable channel regions for ensembles of accumulate-based codes whose transmission takes place over parallel AWGN channels. The simplest ensemble we consider is the ensemble of uniformly interleaved and non-systematic repeat-accumulate (NSRA) codes with $q \geq 3$ repetitions. It is shown in [9, Section 5] that the third condition in Theorem 3 is satisfied for this ensemble, and more explicitly

$$\sum_{h=1}^{D_n} A_h^{[\mathcal{C}(n)]} = O\left(\frac{1}{n}\right)$$

where $D_n = O(\ln(n))$ (so the sequence $\{D_n\}$ tends to infinity logarithmically with n). Based on the calculations of the distance spectrum of this ensemble (see [9, Section 4]), the fourth condition in Theorem 3 is also satisfied. We note that for this ensemble, the asymptotic growth rate of the distance spectrum satisfies

$$r^{[\mathcal{C}]}(0) = 0, \quad \limsup_{\delta \to 0} \frac{r^{[\mathcal{C}]}(\delta)}{\delta} = \left. \frac{d}{d\delta} \, r^{[\mathcal{C}]}(\delta) \right|_{\delta=0} = 0.$$

Hence, inequality (100) in Theorem 3 (i.e., the second condition in this theorem) is also satisfied for this ensemble (since the RHS of (100) is always positive). Hence, the fulfillment of all the conditions in Theorem 3 for this ensemble requires to check under which conditions the error exponent is strictly positive (see the condition in (99)).

As a second example, for the Gallager ensembles of regular (n, j, k) LDPC codes, the second, third and fourth conditions are also satisfied for the case where $j \geq 3$. Under this assumption, the minimum distance even grows linearly with the block length (see [14, Section 2.2]), so the LHS of (100) becomes negative.

We make use of the fulfillment of the condition in (100) for regular NSRA codes and some other variants of accumulate-based codes later in Section VII-B.

It is important to note that the low Hamming weight codewords which are addressed by the requirement in (101) may yield that the error probability under ML decoding does not necessarily vanish exponentially with the block length (see, e.g., [26, Theorems 3 and 4] and [9, Section 5], where the ML decoding error probability of the considered ensembles of turbo-like codes vanish asymptotically like the inverse of a polynomial of the block length).

VII. Performance Bounds for Turbo-Like Ensembles over Parallel Channels: Moderate Block Lengths and the Asymptotic Case

In this section, we exemplify the performance bounds derived in this paper for various ensembles of turbo-like codes whose transmission is assumed to take place over parallel BIAWGN channels. We also compare the bounds to those introduced in [23], showing the superiority of the new bounds introduced in Sections III and IV. As mentioned before, the superiority of the generalized 1961 Gallager bound in Section IV over the LMSF bound from [23] is attributed to the optimizations of the related tilting measures over each of the individual channels.

We especially focus on ensembles of uniformly interleaved repeat-accumulate (RA) codes and accumulate-repeat-accumulate (ARA) codes. These codes, originally introduced by Divsalar et al. [2], [9], are attractive since they possess low encoding and decoding complexity under iterative decoding and remarkable improvement in performance over classical algebraic codes. For independent parallel channels, we study their theoretical performance under ML decoding. The section considers both finite-length analysis and asymptotic analysis. In the former case, we present upper bounds on the ML decoding error probability, and in the latter case, we consider inner bounds on the attainable channel regions of these ensembles and study the gap to the capacity region. In order to assess the tightness of the bounds for ensembles of relatively short block lengths, we compare the upper bounds under optimal ML decoding with computer simulations under (sub-optimal) iterative decoding. In some cases, the upper bounds under ML decoding are more pessimistic than the experimental results of the iterative decoder, thus indicating that there is still room for improving the tightness of the new bounds.

The structure of this section is as follows. Section VII-A exemplifies performance bounds for ensembles of short to moderate block length by focusing on a uniformly interleaved ensemble of turbo codes, comparing various bounds on the bit error probability under ML decoding and compare the results with computer simulation of the Log-MAP iterative decoding. Section VII-B focuses on performance bounds for repeat-accumulate codes and their recent variations which are attractive due to their remarkable performance and low encoding and decoding complexity under iterative decoding algorithms. This sub-section analyzes the performance under optimal ML decoding, assuming the communications takes place over parallel channels. The inner bounds on the attainable channel regions whose calculations are based on Theorem 3 considerably extend the channel region which corresponds to the cutoff rate, and outperform previously reported bounds. We conclude the discussion in this section with practical considerations related to efficient implementations of the generalized DS2 and 1961 Gallager bounds for parallel channels, thus aiming to reduce the computational complexity related to the evaluation of these bounds (see Section VII-C).
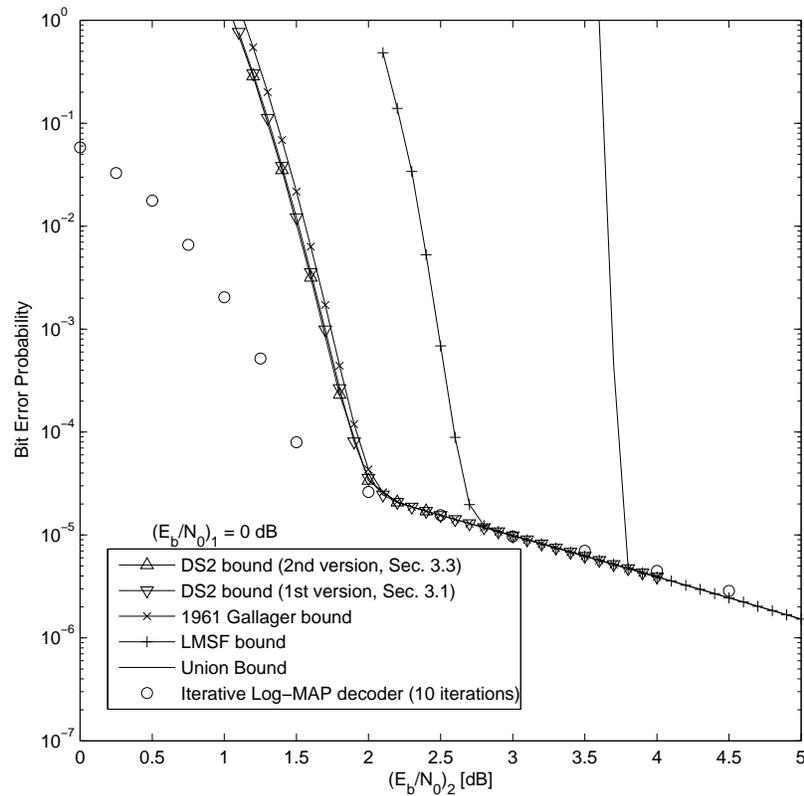
A. Performance Bounds for Uniformly Interleaved Turbo Codes

In this sub-section, we exemplify the tightness of the new bounds by referring to an ensemble of uniformly interleaved turbo codes, and comparing the upper bounds on the bit error probability under ML decoding with computer simulations of an iterative decoder. The bounds are exemplified for the transmission over two parallel BIAWGN channels. The reader is referred to [21] which introduces coding theorems for turbo code ensembles under ML decoding, assuming that the transmission takes place over a single MBIOS channel (i.e., $J = 1$ in our setting).

Fig. 4 compares upper bounds on the bit error probability of the ensemble of uniformly interleaved turbo codes of rate $R = \frac{1}{3}$ bits per channel use (see Fig. 4(a)). The calculation of the average distance spectrum and IOWE of this ensemble is performed by calculating the IOWE of the constituent codes which are recursive systematic convolutional codes (to this end, we rely on the general approach provided in [25] for the calculation of the IOWE of convolutional codes), and finally, the uniform interleaver which is placed between the two constituent codes in Fig. 4(a) enables one to calculate the distance spectrum and the IOWE of this ensemble, based on the IOWE of the constituent codes (see [4]). The transmission of the codes from this ensemble is assumed to take place over two (independent) parallel binary-input AWGN channels where each bit is equally likely to be assigned to one of these channels ($\alpha_1 = \alpha_2 = \frac{1}{2}$), and the value of the energy per bit to spectral noise density of the first channel is fixed to $\left(\frac{E_{\rm b}}{N_0} \right)_1 = 0$ dB. Since for long enough block codes, the union bound is not informative at rates beyond the cutoff rate, one would expect that for the considered ensemble of codes (whose block length is roughly 3000 bits), the union bound becomes useless for values of $\left(\frac{E_{\rm b}}{N_0} \right)_2$ below the value in the RHS of (8) (whose value in this setting is 3.69 dB). This limitation of the union bound is indeed reflected from Fig. 4(b), thus showing how loose
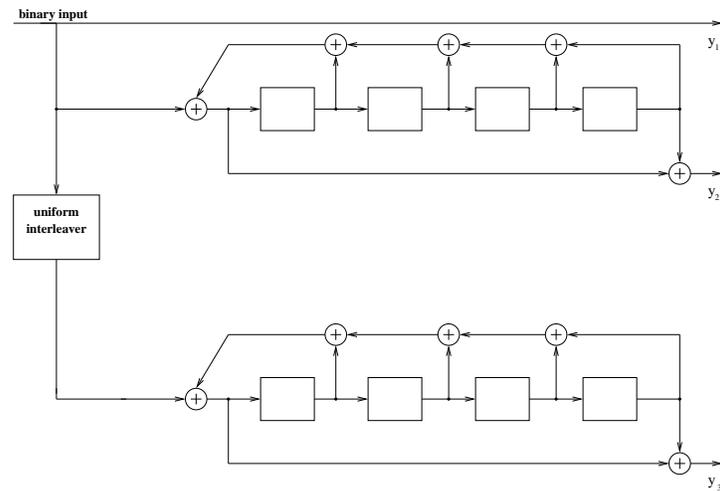
(a) Ensemble of uniformly interleaved turbo codes.

(b) Performance Bounds under ML decoding versus simulation results of iterative Log-MAP decoding.

Fig. 4. (a) The encoder of an ensemble of uniformly interleaved turbo codes whose interleaver is of length 1000 without puncturing of parity bits. (b) Performance bounds for the bit error probability under ML decoding versus computer simulation results of iterative Log-MAP decoding (with 10 iterations). The transmission of this ensemble takes place over two (independent) parallel binary-input AWGN channels. Each bit is equally likely to be assigned to one of these channels, and the energy per bit to spectral noise density of the first channel is set to $\left(\frac{E_b}{N_0}\right)_1 = 0$ dB. The compared upper bounds on the bit error probability are the generalizations of the DS2 and 1961 Gallager bounds, the LMSF bound from [23], and the union bound (based on (C.5)).

is the union bound as compared to computer simulations of the (sub-optimal) iterative decoder. The LMSF bound depicted in Fig. 4(b) uses a partitioning for codes of finite length which was obtained via the algorithm described

in Section V-D; for a bit error probability of 10^{-4} it is about 1 dB tighter than the union bound. Both versions of the DS2 and the 1961 Gallager bounds with their *optimized tilting measures* show a remarkable improvement in their tightness over the union and LMSF bounds where for a bit error probability of 10^{-4}, these three bounds exhibit a gain of 0.8 dB over the LMSF bound. The two versions of the DS2 bound are almost equally tight with a gap between them of less than 0.01dB in favor of the second version. The second version of the DS2 bound gains about 0.05 dB at a bit error probability of 10^{-3} over the 1961 Gallager bound. In spite of a remarkable advantage of the improved bounds over the union and LMSF bounds, computer simulations under (the sub-optimal) iterative Log-MAP decoding with 10 iterations show a gain of about 0.4 dB, so there is still room for further improvement in the tightness of the bounds under ML decoding.

B. Distance Properties and Performance Bounds for Various Ensembles of Accumulate-Based Codes

The IOWEs and the distance spectra of ensembles of irregular repeat-accumulate (IRA) codes and accumulate-repeat-accumulate (ARA) codes were derived in [2], [18]. In the continuation of this section, we compare inner bounds on the attainable channel regions of accumulate-based codes under ML decoding. The comparison refers to three ensembles of rate one-third, as depicted in Fig. 5: the first one is the ensemble of uniformly interleaved and non-systematic RA codes where the number of repetitions is $q = 3$, the second and the third ensembles are uniformly interleaved and systematic ensembles of RA codes and ARA codes, respectively, where the number of repetitions is equal to $q = 6$ and, as a result of puncturing, every third bit of the non-systematic part is transmitted (so the puncturing period is $p = 3$). For simplicity of notation, we make use of the abbreviations NSRA(N, q), SPRA(N, p, q) and SPARA(N, M, p, q) codes for the encoders shown in Figs. 5 (a)–(c) (i.e., the abbreviations 'NS' and 'SP' stand for 'non-systematic' and 'systematic and punctured', respectively). In this notation, N is the input block length.

We rely on the concepts of the analysis introduced in [2] for the calculation of the IOWEs of the uniformly interleaved ensembles in Figs. 5 (a)–(c), as well as the calculation of the asymptotic growth rates of their distance spectra. The generalizations of the DS2 and the 1961 Gallager bounds for parallel channels are then applied to these ensembles for the asymptotic case where we let the block length tend to infinity.

(a) Non-systematic RA codes - NSRA (N, q)

(b) Systematic RA codes with puncturing - SPRA (N, p, q)

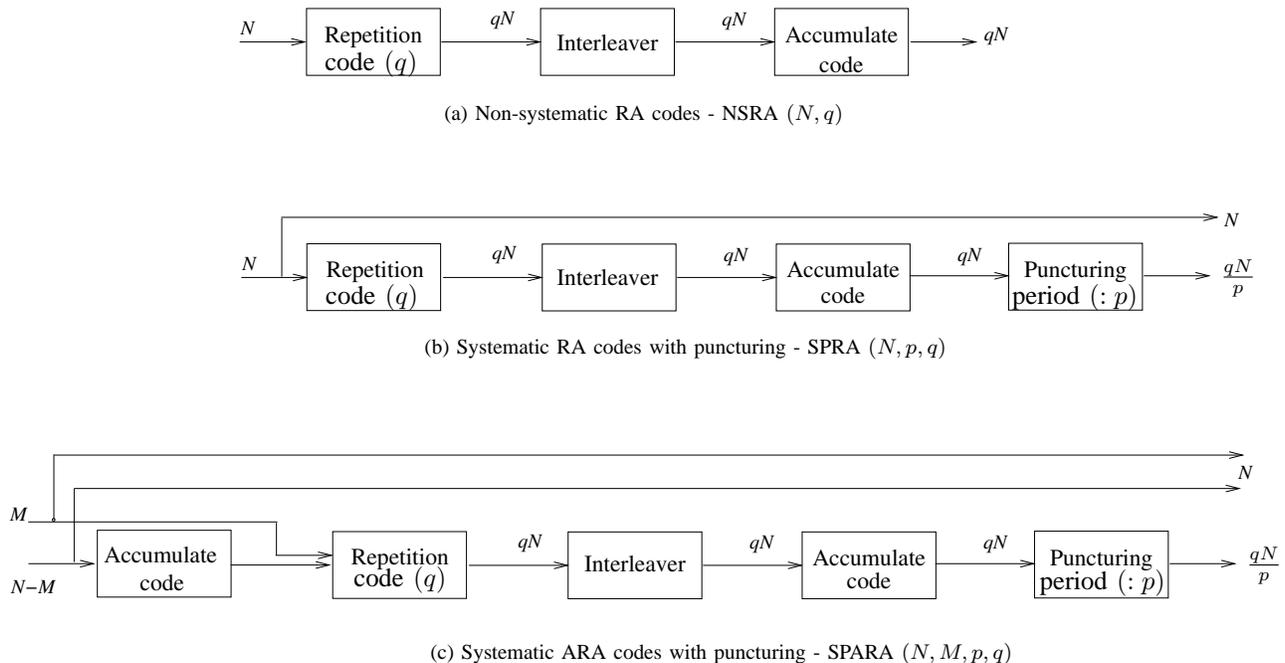(c) Systematic ARA codes with puncturing - SPARA (N, M, p, q)

Fig. 5. Systematic and Non-systematic RA and ARA codes. The interleavers of these ensembles are assumed to be chosen uniformly at random, and are of length qN where N designates the length of the input block (information bits) and q is the number of repetitions. The rates of all the ensembles is set to $\frac{1}{3}$ bits per channel use, so we set $q = 3$ for figure (a), and $q = 6$ and $p = 3$ for figures (b) and (c). The diagram in part (a) corresponds to a non-systematic ensemble without puncturing, whereas the diagrams in parts (b) and (c) correspond to systematic ensembles with puncturing.

The average IOWE and the asymptotic growth rate of the distance spectrum for the ensemble of uniformly interleaved NSRA codes (see Fig. 5 (a)) are given in (16) and (17), respectively.

Following the approach of the analysis in [2], we derive the average IOWEs of the uniformly interleaved ensembles of SPRA(N,3,6) and SPARA(N,M,3,6) codes (see Figs. 5 (b) and (c)). The details of this analysis are provided in Appendix D. In the following, we present the final results related to the finite-length and asymptotic distance properties (where we let N tend to infinity); these results serve later for the calculation of attainable channel regions under ML decoding.

We consider here the case where the repetition code repeats each bit six times, and the parity bits are punctured so that every third bit is transmitted. The IOWE of the ensemble of uniformly interleaved SPRA($N, 3, 6$) codes in Fig. 5 (b) is given by (see Appendix D.1)

$$A_{w,d} = \frac{\binom{N}{w}}{\binom{6N}{6w}} \sum_{h=0}^{2N} \sum_{j=0}^{2N} \sum_{i=\max(0,j-2N+h)}^{\min(j,h)} \left\{ \binom{h}{i} \binom{2N-h}{j-i} \binom{2N-d+w}{\lfloor \frac{h}{2} \rfloor} \right.$$
$$\left. \binom{d-w-1}{\lceil \frac{h}{2} \rceil - 1} 3^{h+j-2i} \, \delta_{6w,2j+h} \right\} \tag{102}$$

where $\delta_{n,m}$ designates the discrete delta function (i.e., it is equal to 1 if $n = m$, and is equal to zero otherwise). The IOWE of the ensemble of uniformly interleaved SPARA($N, M, 3, 6$) codes in Fig. 5 (c) is given by (see Appendix D.1)

$$A_{w,d} = \sum_{m=0}^{M} \sum_{l=0}^{N} \sum_{h=0}^{2N} \sum_{j=0}^{2N} \sum_{i=\max(0,j-2N+h)}^{\min(j,h)} \left\{ \frac{\binom{M}{m} \binom{N-M-l+m}{\lfloor \frac{w-m}{2} \rfloor} \binom{l-m-1}{\lceil \frac{w-m}{2} \rceil - 1}}{\binom{6N}{6l}} \right.$$
$$\binom{h}{i} \binom{2N-h}{j-i} \binom{2N-d+w}{\lfloor \frac{h}{2} \rfloor}$$
$$\left. \binom{d-w-1}{\lceil \frac{h}{2} \rceil - 1} 3^{h+j-2i} \, \delta_{6l,2j+h} \right\}. \tag{103}$$

The asymptotic growth rates of the distance spectra of these two ensembles are obtained by the calculation of the limit

$$r^{[\mathcal{C}]}(\delta) = \lim_{N \to \infty} \frac{1}{3N} \sum_{w=0}^{N} A_{w,d}, \qquad \delta = \frac{d}{3N} \ (0 \le \delta \le 1)$$

where it is taken into account that the common rate of these ensembles is equal to one-third bits per channel use.

Let $H(x) = -x \ln(x) - (1-x) \ln(1-x)$ be the binary entropy function to the natural base, then it is well known that

$$\lim_{n \to \infty} \frac{1}{n} \ln \binom{n}{\beta n} = H(\beta), \quad 0 \le \beta \le 1. \tag{104}$$

For the derivation of the asymptotic growth rate of the average distance spectrum for the ensemble of uniformly interleaved SPRA $(N, 3, 6)$ codes where $N \to \infty$, we rely on the IOWE given in (102), and introduce the parameters

$$\eta \triangleq \frac{h}{3N}, \quad \rho_1 \triangleq \frac{i}{3N}, \quad \rho_2 \triangleq \frac{j}{3N}.$$

The asymptotic growth rate of the distance spectrum of this ensemble is given by (see Appendix D.2)

$$r(\delta) = \max_{\eta,\rho_1,\rho_2} \left\{ -\frac{5}{3} H\left(\frac{2\rho_2 + \eta}{2}\right) + \eta H\left(\frac{\rho_1}{\eta}\right) + \left(\frac{2}{3} - \eta\right) H\left(\frac{\rho_2 - \rho_1}{\frac{2}{3} - \eta}\right) \right.$$
$$+ \left(\frac{2}{3} - \delta + \frac{2\rho_2 + \eta}{6}\right) H\left(\frac{\eta}{2\left(\frac{2}{3} - \delta + \frac{2\rho_2 + \eta}{6}\right)}\right)$$
$$\left. + \left(\delta - \frac{2\rho_2 + \eta}{6}\right) H\left(\frac{\eta}{2\left(\delta - \frac{2\rho_2 + \eta}{6}\right)}\right) + (\eta + \rho_2 - 2\rho_1) \ln 3 \right\} \tag{105}$$

where the three-parameter maximization is performed over the finite domain which is characterized by the following inequalities:

$$0 \leq \eta \leq \frac{2}{3}, \quad 0 \leq \rho_2 \leq \frac{2}{3}, \quad 2\rho_2 + \eta \leq 6\delta, \quad \rho_2 + 2\eta \leq 3\delta$$

$$\max\left(0, \rho_2 + \eta - \frac{2}{3}\right) \leq \rho_1 \leq \min(\rho_2, \eta), \quad \eta - \rho_2 + 3\delta \leq 2. \tag{106}$$

The derivation of the asymptotic growth rate of the average distance spectrum for the ensemble of uniformly interleaved SPARA $(N, M, 3, 6)$ codes, where $N \to \infty$ and the ratio $\frac{M}{N}$ is fixed, relies on the IOWE given in (103). To this end, we introduce the three additional parameters

$$\alpha \triangleq \frac{M}{3N}, \quad \varepsilon_1 \triangleq \frac{m}{3N}, \quad \varepsilon_2 \triangleq \frac{w-m}{3N}. \tag{107}$$

As mentioned above, the value of α is fixed, and also $0 \leq \alpha \leq \frac{1}{3}$ (since $M \leq N$). After straightforward and tedious algebra which is conceptually similar to the calculations in Appendix D.2, one obtains the following expression for the asymptotic growth rate of the average distance spectrum of the considered ensemble of uniformly interleaved SPARA codes:

$$r(\delta) = \max_{\eta, \rho_1, \rho_2, \varepsilon_1, \varepsilon_2} \left\{ \alpha \, H\left(\frac{\varepsilon_1}{\alpha}\right) + \left(\frac{1}{3} - \alpha - \frac{2\rho_2 + \eta}{6} + \varepsilon_1\right) H\left(\frac{\varepsilon_2}{2\left(\frac{1}{3} - \alpha - \frac{2\rho_2 + \eta}{6} + \varepsilon_1\right)}\right) \right.$$

$$+ \left(\frac{2\rho_2 + \eta}{6} - \varepsilon_1\right) H\left(\frac{\varepsilon_2}{2\left(\frac{2\rho_2 + \eta}{6} - \varepsilon_1\right)}\right) + \eta \, H\left(\frac{\rho_1}{\eta}\right) - 2H\left(\frac{2\rho_2 + \eta}{2}\right)$$

$$+ \left(\frac{2}{3} - \eta\right) H\left(\frac{\rho_2 - \rho_1}{\frac{2}{3} - \eta}\right) + \left(\frac{2}{3} - \delta + \varepsilon_1 + \varepsilon_2\right) H\left(\frac{\eta}{2\left(\frac{2}{3} - \delta + \varepsilon_1 + \varepsilon_2\right)}\right)$$

$$\left. + (\delta - \varepsilon_1 - \varepsilon_2) H\left(\frac{\eta}{2(\delta - \varepsilon_1 - \varepsilon_2)}\right) + (\eta + \rho_2 - 2\rho_1) \ln 3 \right\} \tag{108}$$

where the five-parameter maximization is performed over the finite domain which is characterized by the following inequalities:

$$0 \leq \eta \leq \frac{2}{3}, \quad 0 \leq \rho_2 \leq \frac{2}{3}, \quad 0 \leq \varepsilon_1 \leq \alpha,$$

$$0 \leq \varepsilon_1 + \varepsilon_2 \leq \min\left(\delta, \frac{1}{3}\right),$$

$$\max\left(0, \rho_2 + \eta - \frac{2}{3}\right) \leq \rho_1 \leq \min(\rho_2, \eta),$$

$$0 \leq \eta \leq \min\left(\frac{4}{3} - 2\delta + 2(\varepsilon_1 + \varepsilon_2), 2\delta - 2(\varepsilon_1 + \varepsilon_2)\right),$$

$$\varepsilon_2 \leq \min\left(\frac{2}{3} - 2\alpha - \frac{2\rho_2 + \eta}{3} + 2\varepsilon_1, \frac{2\rho_2 + \eta}{3} - 2\varepsilon_1\right). \tag{109}$$

In the asymptotic case where we let the block length tend to infinity, inner bounds on the attainable channel regions for the considered ensembles of accumulate-based codes are calculated in this section by Theorem 3.

In Fig. 6, we compare inner bounds on the attainable channel boundaries as calculated by the union, LMSF, and DS2 bounds from Sections III-A and III-B. This plot refers to the ensemble of NSRA$(N, 3)$ codes of rate $\frac{1}{3}$ bits per channel use (see Fig. 5 (a)) where we let N tend to infinity. The asymptotic growth rate of the distance spectrum of this ensemble is calculated by (17) with $q = 3$. The remarkable superiority of the both versions of the DS2 bound over the union and LMSF bounds is exemplified for this ensemble of turbo-like codes; actually, the DS2 bound from Sec. III-B appears to be slightly tighter than the DS2 bound from Sec. III-A at the extremities of the boundary of the attainable channel region. We conjecture that this is the region where the application of Jensen's inequality in the latter bound (see the move from (28) to (29)) hinders its tightness the most, possibly due to the large variance of the summands in (28). This phenomenon was also observed for various turbo-like ensembles,
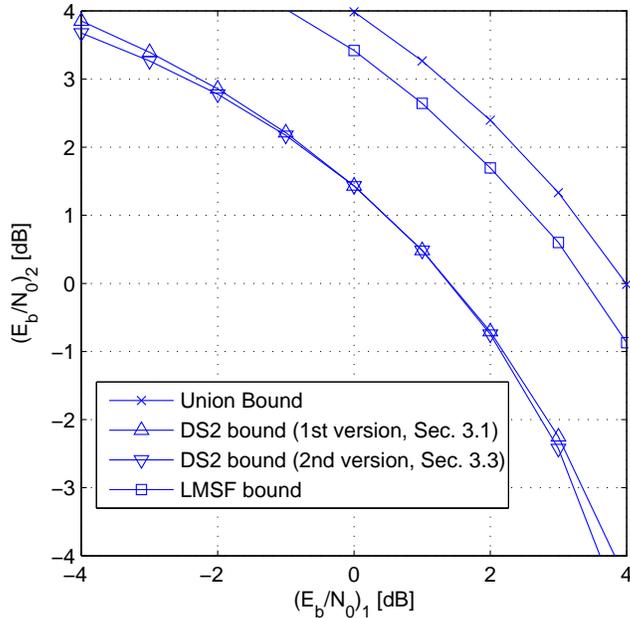
Fig. 6.   Attainable channel regions for the rate one-third uniformly interleaved ensemble of NSRA$(N, 3)$ codes (see Fig. 5 (a)) in the asymptotic case where we let N tend to infinity. The communication takes place over $J = 2$ parallel binary-input AWGN channels, and the bits are equally likely to be assigned over one of these channels ($\alpha_1 = \alpha_2 = \frac{1}{2}$). The achievable channel region refers to optimal ML decoding. The boundaries of the union and LMSF bounds refer to the discussion in [23], while the boundaries referring to the two versions of the DS2 bound refer to the derivations in Sections III-A and III-B, followed by an optimization of the tilting measures derived in these sections.

as well as for ensembles of fully random block codes. However, in the middle region where the channels are not very different, the DS2 bound from Sec. III-A is in some cases tighter than the DS2 bound from Sec. III-B. In the continuation of this section, we therefore compare inner bounds on the attainable channel regions for various ensembles of turbo-like codes where the boundaries of these regions by choosing the tightest version of the DS2 bound, i.e., that which yields the largest attainable channel region. This comparison appears in Fig. 8).

Fig. 7 compares the asymptotic growth rate of the distance spectra of several ensembles of uniformly interleaved and accumulate-based codes where the ensembles are depicted in Fig. 5. The improved performance of the ensembles of SPARA codes under ML decoding is demonstrated in Fig. 8. This improvement is attributed to the distance spectral thinning effect [27] which is exemplified in Fig. 7 for the ensembles of NSRA, SPRA and SPARA codes of the same code rate ($\frac{1}{3}$ bits per channel use). The same phenomenon of distance spectral thinning also exists by reducing the value of α for the ensembles of SPARA codes (see Fig. 7, comparing the two plots for $\alpha = \frac{1}{4}$ and $\alpha = \frac{2}{15}$); this in turn yields an improved inner bound on the attainable channel regions, as observed in Fig. 8. It is shown in this figure that for the SPARA ensemble with the parameters $p = 3, q = 6$ and $\alpha = \frac{2}{15}$, the gap between the inner bound on the attainable channel region under ML decoding and the capacity limit is less than 0.05 dB. Note that for the examined ensembles of NSRA and SPRA codes of the same code rate, the corresponding gaps between the inner bounds on the attainable channel regions and the channel capacity are 2.2 dB and 0.5 dB, respectively (see Fig. 8).

C. Considerations on the Computational Complexity of the Generalized DS2 and 1961 Gallager Bounds

The brute-force calculation of the generalized DS2 bound for linear codes (or ensembles) of finite length is in general computationally heavy. For every constant weight subcode, it requires a numerical optimization over the two parameters $\lambda \geq 0$ and $0 \leq \rho \leq 1$; for each subcode of constant Hamming weight and for each choice of values for λ and ρ, one needs to solve numerically the implicit equations for k and β_j (see Eqs. (43) and (44)) which are related to the J optimized tilting measures. Moreover, for each subcode and a pair of values for λ and ρ, the evaluation of the generalized DS2 bound requires numerical integrations (or summations, in case the channel
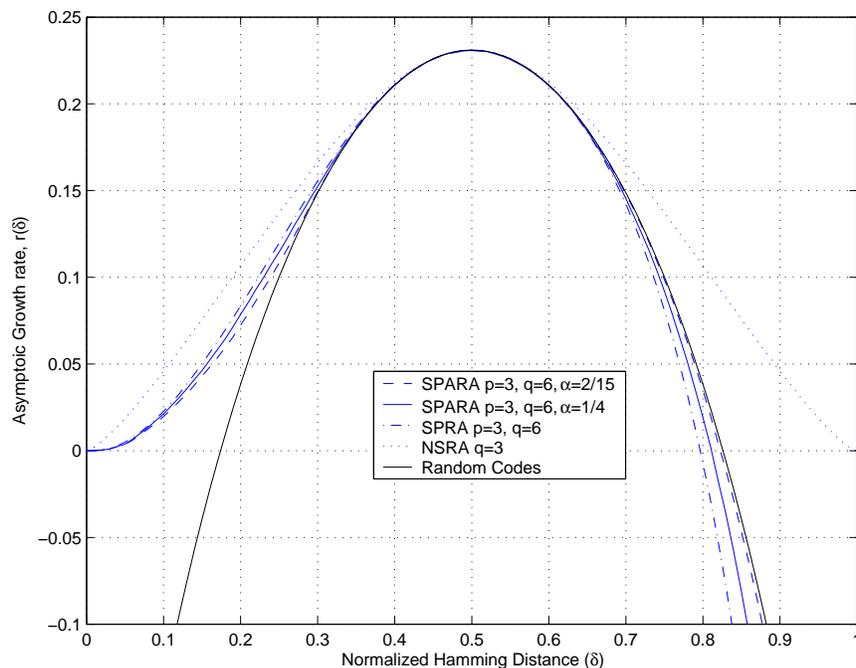
Fig. 7. Comparison of asymptotic growth rates of the average distance spectra of ensembles of RA and ARA codes.

outputs are discrete). Performing these tedious and time-consuming optimizations for every constant weight subcode would make the improved bounds less attractive in terms of their practical use for performance evaluation of linear codes and ensembles.

In the following, we suggest an approach which significantly reduces the complexity related to the computation of the generalized DS2 bound, and enhances the applicability of the bound using standard computational facilities. First, the code is partitioned into constant Hamming weight subcodes, and the exact union bound (see Eq. (C.5)) is calculated for every subcode (note that the number of subcodes does not exceed the block length of the code). This task is rather easy, given the (average) distance spectrum $\{A_h\}$ or the weighted IOWE $\{A'_h\}$ of the code (or ensemble) which are calculated in advance (see (23) and (24)). In order to reduce the computational complexity, we do not calculate the generalized DS2 bounds for those constant-Hamming weight subcodes for which the values of the union bounds fall below a certain threshold (e.g., we may choose a threshold of 10^{-10} for bit error probability or 10^{-6} for block error probability; these thresholds should be tailored for the application under consideration). Next, for those constant Hamming weight subcodes for which the union bound exceeds the above threshold, the generalized DS2 bound is evaluated. For these subcodes, we wish to reduce the infinite interval $\lambda \geq 0$ to a finite interval; this is performed by using the transformation $\lambda' \triangleq \frac{\lambda}{\lambda+1}$ so that the two-parameter optimization is reduced to a numerical optimization over the unit square $(\lambda', \rho) \in [0, 1]^2$. In this respect, it was observed that the optimal values of λ' and ρ vary rather slowly for consecutive values of the constant Hamming weight h, so the search interval associated with the optimization process may be reduced once again with no penalty in the tightness of the bound. In other words, we search for optimal values of λ' and ρ only within a neighborhood of the optimal λ' and ρ found for the previous subcode. We proceed in this manner until all the relevant subcodes are considered. As an example, we note that for the ensemble of turbo codes depicted in Fig. 4(b), about 80% of the computational time was saved without affecting the numerical results; in this respect, the threshold for the bit error probability analysis was chosen to be $\frac{10^{-6}}{n}$ where n designates the block length of the code. The reduction in the computational complexity becomes however more pronounced for higher SNR values, as the number of subcodes for which the union bound replaces the computation of the generalized DS2 bound increases.

An analogous consideration applies to the generalized version of the 1961 Gallager bound for parallel channels with its related optimized tilting measures.

Referring to the calculation of attainable channel regions, a search over the region of channel parameters is required. As an example, consider a set of parallel AWGN channels characterized by the J-tuple of SNRs
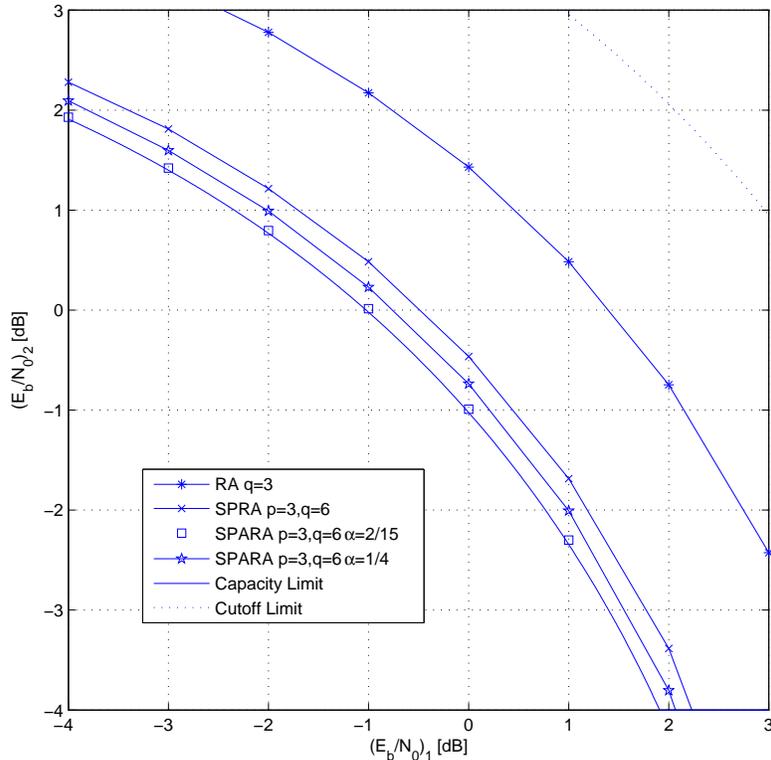
Fig. 8. Attainable channel regions for the rate one-third uniformly interleaved accumulate-based ensembles with puncturing depicted in Fig. 5. These regions refer to the asymptotic case where we let N tend to infinity. The communication takes place over $J = 2$ parallel binary-input AWGN channels, and the bits are equally likely to be assigned over one of these channels ($\alpha_1 = \alpha_2 = \frac{1}{2}$). The achievable channel region refers to optimal ML decoding. The boundaries of these regions are calculated by selecting the tighter of the two generalizations of the DS2 bound appearing in Sections III-A and III-B, followed by the optimization of their respective tilting measures. The capacity limit and the attainable channel regions which corresponds to the cutoff rate are given as a reference.

$(\nu_1, \ldots, \nu_J)$. In order to find the attainable channel boundary, we fix the values of $\nu_1, \ldots, \nu_{J-1}$ and perform a linear search over ν_J using any appropriate method (e.g., the bisection method) in order to find the smallest value of ν_J^* for which the lower bound on the error exponent (as obtained by an upper bound on the ML decoding error probability) vanishes. If $(\nu_1, \ldots, \nu_{J-1}, 0)$ is not an attainable point while $(\nu_1, \ldots, \nu_{J-1}, \infty)$ is attainable, then the resulting value ν_J^* is such that the point $(\nu_1, \ldots, \nu_J^*)$ is on the boundary of the attainable region. The overall complexity of this approach is, of course, polynomial in J. We apply this approach in this section for the calculation of inner bounds on the attainable channel regions under ML decoding, referring to the generalizations of the DS2 and 1961 Gallager bounds in Sections III and IV, respectively.

VIII. SUMMARY

This paper is focused on the performance analysis of binary linear block codes (or ensembles) whose transmission takes place over independent, memoryless and symmetric parallel channels. New bounds on the maximum-likelihood (ML) decoding error probability are derived. These bounds are applied to various ensembles of turbo-like codes, focusing especially on punctured repeat-accumulate and accumulate-repeat-accumulate (ARA) codes which possess low encoding and decoding complexity and exhibit remarkable performance under iterative decoding (see, e.g., [1], [9], [20], [28], [29]). The framework of the second version of the Duman and Salehi (DS2) bounds is generalized to the case of parallel channels by means of two different approaches, along with the derivation of their optimized tilting measures. For the case of random codes, one of the bounds (namely, the one derived in Sec. III-B) attains the random coding exponent while the other (derived in Sec. III-A) does not. This difference is attributed to the

additional Jensen's inequality in the transition from (28) to (29) (see p. 11) which is circumvented in the derivation of Sec. III-B. Nevertheless, for general code ensembles, neither of these two bounds is tighter than the other. The generalization of the 1961 Gallager bound for parallel channels, introduced by Liu at al. [23], is reviewed and the optimized tilting measures which are related to this bound are derived via calculus of variations (as opposed to the use of simple and sub-optimal tilting measures in [23]). The connection between the generalized DS2 bound and the 1961 Gallager bound, which was originally addressed by Divsalar [10] and by Sason and Shamai [33], [36] for a single channel, is revisited in this paper for an arbitrary number of independent parallel channels which are memoryless and symmetric. In this respect, it is shown that the 1961 Gallager bound [23] is a special case of the generalized DS2 bound derived in Sec. III-B and is not a special case of the DS2 bound derived in Sec. III-A. In the asymptotic case where we let the block length tend to infinity, the new bounds are used to obtain improved inner bounds on the attainable channel regions under ML decoding. The tightness of the new bounds for independent parallel channels is exemplified for structured ensembles of turbo-like codes. In this respect, the inner bounds on the attainable channel regions which are computed by the DS2 bound from Sec. III-A are slightly looser than those computed by the DS2 bound from Sec. III-B at the extremities of the boundary of the attainable channel region. On the other hand, in the region where the channels are not very different, the DS2 bound from Sec. III-A is slightly tighter. It is therefore suggested to use in each case the tighter of the two bounds in order to maximize the attainable channel region. For turbo-like ensembles of moderate block lengths, the two versions of the generalized DS2 bound are almost equally tight (see, e.g., Fig. 4(b) in p. 32).

Following the approach in [2], we analyze the distance spectra and their asymptotic growth rates for various ensembles of systematic and punctured accumulate-based codes (see Fig. 5). This distance spectral analysis serves to assess the performance of these codes under ML decoding where we rely on the bounding techniques developed in [23] and this paper for parallel channels. The improved performance of the ensembles of systematic and punctured accumulate-repeat-accumulate (SPARA) codes under ML decoding is demonstrated by combining the two generalized DS2 bounds from Sections III-A and III-B (see Fig. 8). This improvement is attributed to the distance spectral thinning effect [27] which is exemplified in Fig. 7 by comparing the asymptotic growth rates of the distance spectra for the ensembles in Fig. 5 (a)–(c).

The generalization of the DS2 bound for parallel channels enables to re-derive specific bounds which were originally derived by Liu et al. [23] as special cases of the 1961 Gallager bound. However, the improved bounds together with their optimized tilting measures show, regardless of the block length of the codes, an improvement over the bounds derived as special cases of the 1961 Gallager bound; this improvement is especially pronounced for moderate to large block lengths. We note that in some cases, the new bounds under ML decoding happen to be a bit pessimistic as compared to computer simulations of sub-optimal iterative decoding (see, e.g., Fig. 4(b)), thus indicating that there is still room for further improvement of the bounds under ML decoding.

APPENDIX A

A. On the Sub-optimality of Even Tilting Measures in the Gallager Bound

In the following, we derive the functions $f(\cdot; j)$ resulting from the optimal DS2 tilting measures in (42) and demonstrate that they are not even functions. From (32), we get the expression

$$\psi(y; j) = \frac{g(y; j)p(y|0; j)}{\displaystyle\sum_{y'} g(y'; j)p(y'|0; j)} \triangleq c^{-1} \cdot g(y; j)p(y|0; j)$$

for the single-letter connection between the normalized and un-normalized DS2 tilting measures; changing the subject gives

$$g(y; j) = c \cdot \left(\frac{\psi(y; j)}{p(y|0; j)} \right) \ . \tag{A.1}$$

Substituting (42) in (A.1) we obtain the optimal form of the un-normalized tilting measure as

$$g(y;j) = c \cdot \left(1 + k \left(\frac{p(y|1;j)}{p(y|0;j)} \right)^\lambda \right)^\rho \tag{A.2}$$

Next, we substitute (79) in the LHS of (A.2) and manipulate the expression to get

$$f(y;j) = \text{const} \cdot p(y|0;j) \left(1 + k \left(\frac{p(y|1;j)}{p(y|0;j)} \right)^\lambda \right)^{\frac{\rho}{s}}. \tag{A.3}$$

Clearly, this expression does not constitute an even function.

B. Technical Details for Calculus of Variations on (84)

The bound on the decoding error probability for constant Hamming weight codes is given by substituting (83) into (72). Disregarding the multiplicative term $2^{h(\rho)}$, we minimize the expression

$$
\begin{aligned}
U \triangleq {} & A_h \left\{ \sum_{j=1}^{J} \alpha_j \sum_y [p(y|0;j)p(y|1;j)]^{\frac{1-r}{2}} f(y;j)^r \right\}^h \\
& \cdot \left\{ \sum_{j=1}^{J} \frac{\alpha_j}{2} \sum_y \left[p(y|0;j)^{1-r} + p(y|1;j)^{1-r} \right] f(y;j)^r \right\}^{n-h} e^{-nrd} \\
& + \left\{ \sum_{j=1}^{J} \frac{\alpha_j}{2} \sum_y \left[p(y|0;j)^{1-s} + p(y|1;j)^{1-s} \right] f(y;j)^s \right\}^n e^{-nsd}, \\
& \qquad\qquad\qquad r \le 0, \; s \ge 0 \; -\infty < d < \infty. \tag{A.4}
\end{aligned}
$$

Employing calculus of variations, we substitute in (A.4) the following tilting measure

$$f(y;j) = f_0(y;j) + \varepsilon\eta(y;j)$$

where $\eta(\cdot; j)$ is an arbitrary function. Next, we impose the condition that $\frac{\partial U}{\partial \varepsilon}\big|_{\varepsilon=0} = 0$ for all $\eta(\cdot; j)$. The derivative is given by

$$
\begin{aligned}
\frac{\partial U}{\partial \varepsilon}\bigg|_{\varepsilon=0} = \; & A_h e^{-nrd} \Bigg\{ h \left[\sum_{j=1}^{J} \alpha_j \sum_y \left(p(y|0;j)p(y|1;j) \right)^{\frac{1-r}{2}} f_0(y;j)^r \right]^{h-1} \\
& \left[\sum_{j=1}^{J} \alpha_j \sum_y \left(p(y|0;j)p(y|1;j) \right)^{\frac{1-r}{2}} r f_0(y;j)^{r-1} \eta(y;j) \right] \\
& \left[\sum_{j=1}^{J} \frac{\alpha_j}{2} \sum_y \left(p(y|0;j)^{1-r} + p(y|1;j)^{1-r} \right) f_0(y;j)^r \right]^{n-h} \\
& + (n-h) \left[\sum_{j=1}^{J} \frac{\alpha_j}{2} \sum_y \left(p(y|0;j)^{1-r} + p(y|1;j)^{1-r} \right) f_0(y;j)^r \right]^{n-h-1} \\
& \left[\sum_{j=1}^{J} \frac{\alpha_j}{2} \sum_y \left(p(y|0;j)^{1-r} + p(y|1;j)^{1-r} \right) r f_0(y;j)^{r-1} \eta(y;j) \right] \\
& \left[\sum_{j=1}^{J} \alpha_j \sum_y \left(p(y|0;j)p(y|1;j) \right)^{\frac{1-r}{2}} f_0(y;j)^r \right]^{h} \Bigg\} \\
& + e^{-nsd} n \left[\sum_{j=1}^{J} \frac{\alpha_j}{2} \sum_y \left(p(y|0;j)^{1-s} + p(y|1;j)^{1-s} \right) f_0(y;j)^s \right]^{n-1} \\
& \left[\sum_{j=1}^{J} \frac{\alpha_j}{2} \sum_y \left(p(y|0;j)^{1-s} + p(y|1;j)^{1-s} \right) s f_0(y;j)^{s-1} \eta(y;j) \right].
\end{aligned}
\tag{A.5}
$$

Defining the constants

$$
\begin{aligned}
c_1 &\triangleq A_h e^{-nrd} hr \left[\sum_{j=1}^{J} \alpha_j \sum_y \left(p(y|0;j)p(y|1;j) \right)^{\frac{1-r}{2}} f_0(y;j)^r \right]^{h-1} \\
c_2 &\triangleq \left[\sum_{j=1}^{J} \frac{\alpha_j}{2} \sum_y \left(p(y|0;j)^{1-r} + p(y|1;j)^{1-r} \right) f_0(y;j)^r \right]^{n-h} \\
c_3 &\triangleq A_h e^{-nrd} \frac{r(n-h)}{2} \left[\sum_{j=1}^{J} \frac{\alpha_j}{2} \sum_y \left(p(y|0;j)^{1-r} + p(y|1;j)^{1-r} \right) f_0(y;j)^r \right]^{n-h-1} \\
c_4 &\triangleq \left[\sum_{j=1}^{J} \alpha_j \sum_y \left(p(y|0;j)p(y|1;j) \right)^{\frac{1-r}{2}} f_0(y;j)^r \right]^{h} \\
c_5 &\triangleq e^{-nsd} \frac{ns}{2} \left[\sum_{j=1}^{J} \frac{\alpha_j}{2} \sum_y \left(p(y|0;j)^{1-s} + p(y|1;j)^{1-s} \right) f_0(y;j)^s \right]^{n-1}
\end{aligned}
\tag{A.6}
$$

and requiring that the integrand in (A.5) be equal to zero, we get the equivalent condition

$$\sum_{j=1}^{J} \alpha_j \left\{ \left(c_1 c_2 \left[p(y|0;j) p(y|1;j) \right]^{\frac{1-r}{2}} + c_3 c_4 \left[p(y|0;j)^{1-r} + p(y|1;j)^{1-r} \right] \right) f_0(y;j)^{r-1} \right.$$
$$\left. + c_5 \left[p(y|0;j)^{1-s} + p(y|1;j)^{1-s} \right] f_0(y;j)^{s-1} \right\} = 0, \quad \forall y \in \mathcal{Y}.$$

Defining $K_1 \triangleq \frac{c_1 c_2}{c_5}$, $K_2 \triangleq \frac{c_3 c_4}{c_5}$, and dividing both sides by $f_0(y;j)^{r-1}$ implies the condition in (85).

APPENDIX B: PROOF OF THEOREM 3

The concept of the proof of this theorem is similar to the proof introduced in [22, pp. 40–42] for the single channel case, and the proofs of [23, Theorems 2–4] for the scenario of independent parallel channels. The difference in this proof from those mentioned above is the starting point which relies on the generalization of the DS2 bound (see Theorem 1 in Section III-A).

We begin by rewriting the DS2 bound for a specific constant Hamming-weight subcode (35) as

$$P_{e|0}(h) \leq A_h^\rho B_h$$

where

$$B_h \triangleq \left\{ \left(\sum_{j=1}^{J} \alpha_j \sum_y \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1-\lambda\rho}{\rho}} p(y|1;j)^\lambda \right)^\delta \right.$$
$$\left. \left(\sum_{j=1}^{J} \alpha_j \sum_y \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1}{\rho}} \right)^{1-\delta} \right\}^{n\rho}. \tag{B.1}$$

By selecting the optimized tilting measures and optimal values of $\lambda \geq 0$ and $0 \leq \rho \leq 1$, we obtain the optimized bound B_h^{opt}, which is related to the optimal exponent $E^{\text{DS2}_1}(\delta)$ by

$$B_h^{\text{opt}} = e^{-n\left(E^{\text{DS2}_1}(\delta) + \rho r^{[\mathcal{C}]}(\delta)\right)}, \quad \delta \triangleq \frac{h}{n}. \tag{B.2}$$

The upper bound on the ML decoding error probability of the ensemble can be written as

$$P_e^{[\mathcal{C}(n)]} \leq \sum_{h=1}^{n} A_h^\rho B_h^{\text{opt}} \leq \sum_{h=1}^{D_n} A_h + \sum_{h=D_n+1}^{\alpha n} A_h \gamma^h + \sum_{h=\alpha n+1}^{n} A_h^\rho B_h^{\text{opt}} \tag{B.3}$$

for any $\alpha > 0$. This follows since for weights up to D_n, we upper bound B_h^{opt} by 1, and for Hamming weights from $D_n + 1$ up to αn, the DS2 bound is relaxed by selecting $\rho = 1$ and using the union bound (see (88)). Let us examine the behavior of each of the three terms in the RHS of (B.3). As we let n tend to infinity, the first term in the RHS of (B.3) goes to 0 due to the third condition of the theorem.

The second term in the RHS of (B.3) satisfies the equality

$$\sum_{h=D_n+1}^{\alpha n} A_h \gamma^h = \sum_{h=D_n+1}^{\alpha n} e^{h\left(\frac{r^{[\mathcal{C}(n)]}(\delta)}{\delta} + \ln(\gamma)\right)}.$$

We now rely on the second condition of Theorem 1 (see (100)) and the fourth condition of this theorem, where the latter condition is related to the uniform convergence of $r^{\mathcal{C}(n)}(\delta)$ to the asymptotic growth rate $r^{[\mathcal{C}]}(\delta)$ over the interval [0,1]. These two condition imply that for large enough values of n, there exists a positive α which is close enough to zero for which

$$\frac{r^{[\mathcal{C}(n)]}(\delta)}{\delta} + \ln(\gamma) < -\theta_0 < 0, \quad \forall \delta \leq \alpha$$

for some positive θ_0. This implies the inequality

$$\sum_{h=D_n+1}^{\alpha n} e^{h\left(\frac{r^{[\mathcal{C}(n)]}(\delta)}{\delta}+\ln(\gamma)\right)} \leq \sum_{h=D_n+1}^{\alpha n} e^{-h\theta_0} \leq \frac{e^{-D_n\theta_0}}{1-e^{-\theta_0}}$$

which tends to zero as $n \to \infty$ because $D_n \to \infty$.

Finally, by using (B.2) and relying on the fourth condition in Theorem 1, the third term in the RHS of (B.3) can be expressed as

$$\sum_{h=\alpha n}^{n} A_h^\rho B_h^{\mathrm{opt}} = \sum_{h=\alpha n}^{n} e^{-n\left(E^{\mathrm{DS2}_1}(\delta)+\rho\left(r^{[\mathcal{C}]}(\delta)-r^{[\mathcal{C}(n)]}(\delta)\right)\right)} = \sum_{h=\alpha n}^{n} e^{-n\left(E^{\mathrm{DS2}_1}(\delta)+o(1)\right)} \tag{B.4}$$

which vanishes as $n \to \infty$ due to (99), thus completing the proof of Theorem 1 (since it was demonstrated that each of the three terms in the RHS of (B.3) vanishes as we let n tend to infinity, so also the decoding error probability in the LHS of (B.3) vanishes asymptotically). $\square$

APPENDIX C: EXACT UNION BOUND FOR PARALLEL GAUSSIAN CHANNELS

In this appendix, we derive the union bound on the ML decoding error probability of binary linear block codes transmitted over parallel Gaussian channels. This form of the union bound can also be used in conjunction with other bounds (e.g., 1961 Gallager or both versions of the DS2 bounds) for constant Hamming weight subcodes in order to tighten the resulting bound. Unfortunately, we cannot compare here "two versions" of the union derived by the two different approaches which were used for the DS2 bound in Sec. III. This is because when applying the CSIR model, we have no exact expression for the pairwise error probability for a general distribution α_j of the channel states. Therefore, we must use the first approach of averaging the bound over all possible channel mappings. We start the derivation by expressing the pairwise error probability given that the all-zero codeword is transmitted

$$P_{\mathrm{e}}(\underline{0} \to \underline{x}_{h_1,h_2,\ldots,h_J}) = Q\left(\sqrt{2\sum_{j=1}^{J}\nu_j h_j}\right) \tag{C.1}$$

where $\underline{x}_{h_1,h_2,\ldots,h_J}$ is a codeword possessing split Hamming weights $h_1,\ldots,h_J$ in the J parallel channels, and $\nu_j \triangleq \left(\frac{E_s}{N_0}\right)_j$ designates the energy per symbol to spectral noise density for the jth AWGN channel ($j=1,2,\ldots,J$). The union bound on the block error probability gives

$$P_{\mathrm{e}} \leq \sum_{h=1}^{n} \sum_{\substack{h_1\geq 0,\ldots,h_J\geq 0 \\ h_1+\ldots+h_J=h}} A_{h_1,\ldots,h_J} Q\left(\sqrt{2\sum_{j=1}^{J}\nu_j h_j}\right) \tag{C.2}$$

where this bound is expressed in terms of the split weight enumerator of the code. Averaging (C.2) over all possible channel assignments and codes from the ensemble gives (see (30))

$$\overline{P_{\mathrm{e}}} \leq \sum_{\substack{n_j\geq 0 \\ n_1+\ldots+n_J=n}} \left\{\sum_{h=1}^{n} \sum_{\substack{0\leq h_j\leq n_j \\ \sum_j h_j=h}} \overline{A_h}\, P_{\underline{H}|\underline{N}}(\underline{h}|\underline{n})\, P_{\underline{N}}(\underline{n})\, Q\left(\sqrt{2\sum_{j=1}^{J}\nu_j h_j}\right)\right\}$$

$$= \sum_{\substack{n_j\geq 0 \\ n_1+\ldots+n_J=n}} \left\{\sum_{h=1}^{n} \sum_{\substack{0\leq h_j\leq n_j \\ \sum_j h_j=h}} \overline{A_h}\binom{h}{h_1,h_2,\ldots,h_J}\right.$$

$$\binom{n-h}{n_1-h_1,n_2-h_2,\ldots,n_J-h_J}$$

$$\left.\alpha_1^{n_1}\ldots\alpha_J^{n_J} Q\left(\sqrt{2\sum_{j=1}^{J}\nu_j h_j}\right)\right\} \tag{C.3}$$

where α_j designates the a-priori probability for the transmission of symbols over the jth channel, assuming the assignments of these symbols to the J parallel channels are independent and random.

In order to simplify the final result, we rely on Craig's identity for the Q-function [7], i.e.,

$$Q(x) = \frac{1}{\pi} \int_0^{\frac{\pi}{2}} e^{-\frac{x^2}{2\sin^2\theta}} \, \mathrm{d}\theta, \qquad x \geq 0. \tag{C.4}$$

Plugging (C.4) into (C.3) and interchanging the order of integration and summation gives

$$
\begin{aligned}
\overline{P_{\mathrm{e}}} \;\leq\; & \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \sum_{\substack{n_j \geq 0 \\ n_1 + \ldots + n_J = n}} \left\{ \sum_{h=1}^{n} \sum_{\substack{0 \leq h_j \leq n_j \\ \sum h_j = h}} \overline{A_h} \binom{h}{h_1, h_2, \ldots, h_J} \right. \\
& \left. \binom{n-h}{n_1 - h_1, n_2 - h_2, \ldots, n_J - h_J} \alpha_1^{n_1} \cdots \alpha_J^{n_J} \prod_{j=1}^{J} e^{-\frac{\nu_j h_j}{\sin^2\theta}} \right\} \mathrm{d}\theta \\
\overset{\text{(a)}}{=}\; & \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \sum_{h=1}^{n} \overline{A_h} \sum_{\substack{h_j \geq 0 \\ \sum h_j = h}} \left\{ \binom{h}{h_1, h_2, \ldots, h_J} \prod_{j=1}^{J} \left[\alpha_j e^{-\frac{\nu_j}{\sin^2\theta}} \right]^{h_j} \right\} \\
& \sum_{\substack{k_j \geq 0 \\ \sum_j k_j = n-h}} \left\{ \binom{n-h}{k_1, k_2, \ldots, k_J} \prod_{j=1}^{J} (\alpha_j)^{k_j} \right\} \mathrm{d}\theta \\
\overset{\text{(b)}}{=}\; & \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \sum_{h=1}^{n} \overline{A_h} \left[\sum_{j=1}^{J} \alpha_j e^{-\frac{\nu_j}{\sin^2\theta}} \right]^h \mathrm{d}\theta
\end{aligned}
\tag{C.5}
$$

where (a) follows by substituting $k_j = n_j - h_j$ for $j = 1, 2, \ldots, J$, and (b) follows since the sequence $\{\alpha_j\}_{j=1}^{J}$ is a probability distribution, which gives the equality

$$\sum_{\substack{k_j \geq 0 \\ \sum_j k_j = n-h}} \left\{ \binom{n-h}{k_1, k_2, \ldots, k_J} \prod_{j=1}^{J} (\alpha_j)^{k_j} \right\} = \left(\sum_{j=1}^{J} \alpha_j \right)^{n-h} = 1.$$

Eq. (C.5) provides the exact (Q-form) version of the union bound on the block error probability for independent parallel AWGN channels.

APPENDIX D: DISTANCE SPECTRA ANALYSIS OF SYSTEMATIC ACCUMULATE-BASED CODES WITH PUNCTURING

The following analysis is focused on the distance spectra of uniformly interleaved and systematic ensembles of repeat-accumulate (RA) codes and accumulate-repeat-accumulate (ARA) codes with puncturing (see Figs. 5 (b) and (c) in p. 33). As mentioned in Section VII-B, these two ensembles are abbreviated by SPRA and SPARA codes, respectively (where 'SP' stands for 'systematic and punctured'). We derive here the input-output weight enumerator (IOWEs) of these ensembles and also calculate the asymptotic growth rates of their distance spectra. The analysis follows the approach introduced in [2], and it is written in a self-contained manner.

The component codes constructing SPRA and SPARA codes are an accumulate code (i.e., a rate-1 differential encoder), a repetition code and a single parity-check (SPC) code. Since we consider ensembles of uniformly interleaved codes, their IOWEs depend on the IOWE of the above component codes [4], [5]. As a preparatory step, we introduce the IOWEs of these components.

1) The IOWE of a repetition (REP) code is given by

$$A_{w,d}^{\mathrm{REP}(q)} = \binom{k}{w} \delta_{d,qw} \tag{D.1}$$

where k designates the input block length, and $\delta_{n,m}$ is the discrete delta function.

2) The IOWE of an accumulate (ACC) code is given by

$$A_{w,d}^{\text{ACC}} = \binom{n-d}{\lfloor \frac{w}{2} \rfloor}\binom{d-1}{\lceil \frac{w}{2} \rceil - 1}$$ (D.2)

where n is the block length (since this code is of rate 1, the input and output block lengths are the same). The IOWE in (D.2) can be easily obtained combinatorially; to this end, we rely on the fact that for the accumulate code, every single '1' at the input sequence flips the value at the output from this point (until the occurrence of the next '1' at the input sequence).

3) The IOWE function of a non-systematic single parity-check code which provides the parity bit of each set of p consecutive bits, call it SPC(p), is given by (see [2, Eq. (8)])

$$\begin{aligned} A(W,D) &= \sum_{w=0}^{np}\sum_{d=0}^{n} A_{w,d}^{\text{SPC}(p)} W^w D^d \\ &= \Big[\text{Even}\big((1+W)^p\big) + \text{Odd}\big((1+W)^p\big)D\Big]^n \end{aligned}$$ (D.3)

where

$$\begin{aligned} \text{Even}\big((1+W)^p\big) &= \frac{(1+W)^p + (1-W)^p}{2} \\ \text{Odd}\big((1+W)^p\big) &= \frac{(1+W)^p - (1-W)^p}{2} \end{aligned}$$ (D.4)

are two polynomials which include the terms with the even and odd powers of W, respectively.

To verify (D.3), note that a parity-bit of this code is equal to 1 if and only if the number of ones in the corresponding set of p bits is odd; also, the number of check nodes in the considered code is equal to the block length of the code (n).
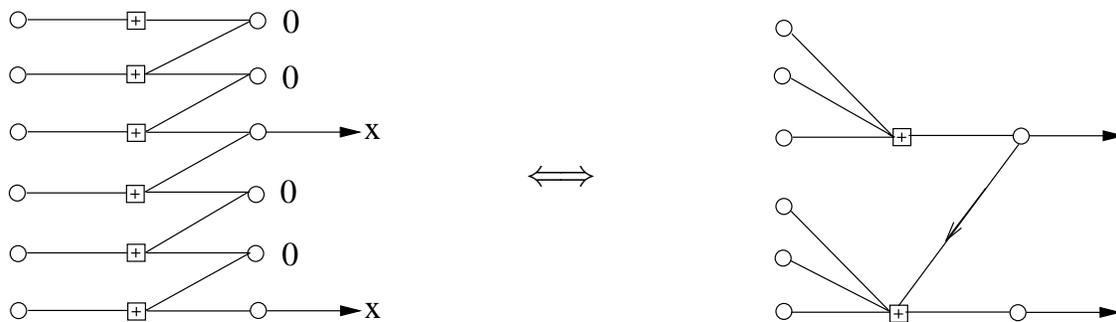

Fig. 9.   Accumulate code with puncturing period $p = 3$ and an equivalent version of an SPC(p) code followed by an accumulate code.

The case where the output bits of an accumulate code are punctured with a puncturing period p is equivalent to an SPC(p) code followed by an accumulate code (see Fig. 9 which was originally shown in [2, Fig. 2]). Hence, for the uniformly interleaved ensembles of SPRA and SPARA codes with a puncturing period of $p = 3$ (see Figs. 5 (b) and (c)), we are interested in the IOWE of the SPC(3) code. For the case where $p = 3$, (D.4) gives

$$\text{Even}\big((1+W)^3\big) = 1 + 3W^2, \quad \text{Odd}\big((1+W)^3\big) = 3W + W^3$$

and (D.3) gives straightforwardly the following IOWE of the SPC(3) code [2, Eq. (15)]:

$$A_{w,d}^{\text{SPC}(3)} = \binom{n}{d}\sum_{j=0}^{n}\sum_{i=\max(0,j-n+d)}^{\min(j,d)}\binom{d}{i}\binom{n-d}{j-i}3^{d+j-2i}\,\delta_{w,2j+d}.$$ (D.5)

In the following, we consider two uniformly interleaved ensembles of SPRA and SPARA codes with $q = 6$ repetitions and puncturing of period $p = 3$, as shown in Fig. 5 (b) and (c). We rely here on the equivalence shown

in Fig. 9, related to the inner accumulate code with puncturing. In this respect, since the input bits to the SPC (appearing in the right plot in Fig. 9) are permuted by the uniform interleaver which is placed after the repetition code (see Figs. 5 (b) and (c)), then the average IOWEs of these two ensembles remain unaffected by placing an additional uniform interleaver between the SPC and the inner accumulate codes. Similarly, placing another uniform interleaver between the precoder in Fig. 5 (c) (i.e., the code which accumulates the first $N - M$ bits) and the repetition code, does not affect the average IOWE of the overall ensemble in Fig. 5 (c).

As mentioned above, the equivalence in Fig. 9 yields that without loss of generality, an additional uniform interleaver of length $N' = \frac{qN}{p} = 2N$ bits can be placed between the SPC(3) code and the accumulate code without affecting the calculation. By doing so, the average IOWE of the serially concatenated and uniformly interleaved ensemble whose constituent codes are the SPC(3) and the accumulate codes, call it ACC(3), is given by (see [5])

$$A_{w,d}^{\text{ACC}(3)} = \sum_{h=0}^{2N} \frac{A_{w,h}^{\text{SPC}(3)} A_{h,d}^{\text{ACC}}}{\binom{2N}{h}}. \tag{D.6}$$

The substitution of (D.2) and (D.5) into (D.6) gives

$$A_{w,d}^{\text{ACC}(3)} = \sum_{h=0}^{2N} \sum_{j=0}^{2N} \sum_{i=\max(0,j-2N+h)}^{\min(j,h)} \left\{ \binom{h}{i} \binom{2N-h}{j-i} \binom{2N-d}{\lfloor \frac{h}{2} \rfloor} \binom{d-1}{\lceil \frac{h}{2} \rceil - 1} \right.$$
$$\left. 3^{h+j-2i} \, \delta_{w,2j+h} \right\}. \tag{D.7}$$

Note that (D.7) is similar to [2, Eq. (19)], except that N in the latter equation is replaced by $2N$ in (D.7). This follows since $\frac{q}{p}$ (i.e., the ratio between the number of repetitions and the puncturing period) is equal here to 2, instead of 1 as was the case in [2]. Equation (D.7) will be used for the finite-length analysis of the distance spectra for the ensembles considered in the continuation to this appendix.

A. Finite-Length Analysis of the Distance Spectra for Systematic Ensembles of RA and ARA Codes with Puncturing

Uniformly Interleaved SPRA $(N, 3, 6)$ *codes*: Let us consider the ensemble depicted in Fig. 5 (b) where $q = 6$ and $p = 3$. Since there is a uniform interleaver of length $N'' = qN$ between the repetition code and the equivalent ACC(3) code, the average IOWE of this serially concatenated and uniformly interleaved systematic ensemble is given by

$$A_{w,d}^{\text{SPRA}(N,3,6)} = \sum_{l=0}^{6N} \frac{A_{w,l}^{\text{REP}(6)} A_{l,d-w}^{\text{ACC}(3)}}{\binom{6N}{l}}$$
$$= \frac{\binom{N}{w} A_{6w,d-w}^{\text{ACC}(3)}}{\binom{6N}{6w}} \tag{D.8}$$

where the last equality is due to the equality in (D.1). Substituting (D.7) in the RHS of (D.8) gives the average IOWE of the considered ensemble, and this result coincides with (102).

Uniformly Interleaved SPARA $(N, M, 3, 6)$ *codes*: By comparing Figs. 5 (b) and (c), we see that a precoder is placed in the second figure. Referring to the ensemble of SPARA codes which is shown in Fig. 5 (c), the precoder is a binary linear block code whose first $N - M$ input bits are accumulated and the other M input bits not changed; these N bits are encoded by the repetition code. The IOWE of the systematic precoder, call it $\text{Pre}(N, M)$, is given by

$$A_{w,d}^{\text{Pre}(N,M)} = \sum_{m=0}^{M} \binom{M}{m} A_{w-m,d-m}^{\text{ACC}}$$
$$= \sum_{m=0}^{M} \left\{ \binom{M}{m} \binom{N-M-d+m}{\lfloor \frac{w-m}{2} \rfloor} \binom{d-m-1}{\lceil \frac{w-m}{2} \rceil - 1} \right\} \tag{D.9}$$

where the last equality relies on (D.2). As mentioned before, for the uniformly interleaved SPARA ensemble depicted in Fig. 5 (c), an additional uniform interleaver between the precoder and the following stages of its encoder does not affect the average IOWE; this ensemble can be therefore viewed as a serial concatenation with a uniform interleaver of length N which is placed between the precoder and the repetition code in Fig. 5 (c) (in addition to the uniform interleaver which is placed after the repetition code). Moreover, referring to the systematic ensemble whose components are REP(6) and ACC(3), the input bits (which are the bits provided by the precoder to the second stage in Fig. 5 (c)) are not transmitted to the channel. In light of these two observations, the average IOWE of the uniformly interleaved ensemble of SPARA codes shown in Fig. 5 (c) is given by

$$A_{w,d}^{\text{SPARA}(N,M,3,6)} = \sum_{l=0}^{N} \frac{A_{w,l}^{\text{Pre}(N,M)} \, A_{l,d-w+l}^{\text{SPRA}(N,3,6)}}{\binom{N}{l}} . \tag{D.10}$$

By substituting (D.8) (i.e., the equality in (102)) and (D.9) into (D.10), one obtains the expression in (103) for the average IOWE of the SPARA$(N, M, 3, 6)$ codes.

B. Asymptotic Analysis of the Distance Spectra

This subsection considers the calculation of the asymptotic growth rates of the distance spectra for the two ensembles in Figs. 5 (b) and (c). The calculation of the asymptotic growth rate of the distance spectrum of a sequence of codes (or ensembles) is performed via (14). In the following, we exemplify the derivation of (105) from the average IOWE in (102). The derivation of (108) from (103) is conceptually similar, but is more tedious algebraically. Since we focus here on ensembles of rate one-third and the block length of the input bits is N (see Fig. 5), the asymptotic growth rate of their distance spectra is obtained by normalizing the logarithm of the average distance spectra of the considered ensemble by $n = 3N$ and letting N tend to infinity. Referring to the average IOWE of the uniformly interleaved ensemble of SPRA$(N, 3, 6)$ codes, as given in (102), we introduce the normalized parameters

$$\delta \triangleq \frac{d}{3N}, \quad \eta \triangleq \frac{h}{3N}, \quad \rho_1 \triangleq \frac{i}{3N}, \quad \rho_2 \triangleq \frac{j}{3N}. \tag{D.11}$$

The normalization by $3N$ yields that the new parameters satisfy

$$0 \le \delta \le 1, \quad 0 \le \eta \le \frac{2}{3}, \quad 0 \le \rho_2 \le \frac{2}{3}. \tag{D.12}$$

From the partial sum w.r.t. the index i in the RHS of (102), dividing the terms in the inequality

$$\max(0, j - 2N + h) \le i \le \min(j, h)$$

by $3N$ gives

$$\max\left(0, \rho_2 + \eta - \frac{2}{3}\right) \le \rho_1 \le \min(\rho_2, \eta). \tag{D.13}$$

Since the codes are systematic and the block length of the input bits is N, we get that the terms which contribute to the IOWE in the RHS of (102) satisfy

$$w \le \min(d, N), \quad 6w = 2j + h \tag{D.14}$$

and, from (D.11), multiplying (D.14) by $\frac{1}{3N}$ gives

$$\frac{2\rho_2 + \eta}{6} \le \min\left(\delta, \frac{1}{3}\right). \tag{D.15}$$

From the binomial coefficients which appear in the RHS of (102), it is required that

$$2N - d + w \ge \left\lfloor \frac{h}{2} \right\rfloor, \quad d - w \ge \left\lceil \frac{h}{2} \right\rceil$$

so dividing both sides of these inequalities by $3N$, and letting N tend to infinity gives

$$\eta - \rho_2 + 3\delta \le 2, \quad \rho_2 + 2\eta \le 3\delta. \tag{D.16}$$

Combining (D.12)–(D.16) gives the domain for the three parameters η, ρ_1 and ρ_2 in (106).

A marginalization of the IOWE enables one to obtain the distance spectrum

$$A_d = \sum_{w=0}^{N} A_{w,d} \tag{D.17}$$

where the IOWE $\{A_{w,d}\}$ is given by (102). Note that unless

$$\frac{w}{N} = \frac{2j+h}{6N} = \frac{2\rho_2 + \eta}{2} \tag{D.18}$$

the IOWE $A_{w,d}$ in (102) vanishes, and therefore it does not affect the sum in the RHS of (D.17). In the limit where $N \to \infty$, the asymptotic growth rate of the average distance spectrum for the uniformly interleaved ensemble of SPRA$(N, 3, 6)$ codes (see Fig. 5 (b) in p. 33) is obtained from (102), (104), and (D.17). Hence, we get

$$
\begin{aligned}
r(\delta) &= \lim_{N \to \infty} \frac{1}{3N} \ln \sum_{w=0}^{N} A_{w,d} \\
&= \lim_{N \to \infty} \max_{h,i,j} \left\{ \frac{1}{3N} \left[NH\left(\frac{w}{N}\right) - 6NH\left(\frac{6w}{6N}\right) + hH\left(\frac{i}{h}\right) + (2N-h)H\left(\frac{j-i}{2N-h}\right) \right. \right. \\
&\qquad\qquad + (2N-d+w)H\left(\frac{h}{2(2N-d+w)}\right) + (d-w-1)H\left(\frac{\frac{h}{2}-1}{d-w-1}\right) \\
&\qquad\qquad \left. \left. + (h+j-2i)\ln 3 \right] \right\}.
\end{aligned}
$$

By multiplying the three parameters which are involved in the maximization by $\frac{1}{3N}$, using (D.11) and (D.18) and taking the limit where N tends to infinity, one readily obtains the result in (105).

REFERENCES

[1] A. Abbasfar, D. Divsalar, and K. Yao, "Accumulate-repeat-accumulate codes," *IEEE Trans. on Communications*, vol. 55, no. 4, pp. 692–702, April 2007.

[2] A. Abbasfar, K. Yao and D. Divsalar, "Maximum-likelihood decoding analysis of accumulate-repeat-accumulate codes," *Proceedings IEEE 2004 Global Telecommunications Conference (GLOBECOM 2004)*, pp. 514–519, 29 November–3 December, 2004, Dallas, Texas, USA.

[3] E. Agrell, "Voronoi regions for binary linear block codes," *IEEE Trans. on Information Theory*, vol. 42, no. 1, pp. 310–316, January 1996.

[4] S. Benedetto and G. Montorsi, "Unveiling turbo codes: some results on parallel concatenated coding schemes," *IEEE Trans. on Information Theory*, vol. 42, no. 2, pp. 409–429, March 1996.

[5] S. Benedetto, D. Divsalar, G. Montorsi and F. Pollara, "Serial concatenation of interleaved codes: Performance analysis, design and iterative decoding," *IEEE Trans. on Information Theory*, vol. 44, no. 3, pp. 909–926, May 1998.

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

[7] J. W. Craig, "A new, simple and exact result for calculating the probability of error for two-dimensional signal constellations," *Proceedings 1991 IEEE Global Communications*, pp. 25.5.1–25.5.5, Boston, MA, USA, 1991.

[8] C. Di, T. Richardson and R. Urbanke, "Weight distribution of low-density parity-check codes," *IEEE Trans. on Information Theory*, vol. 52, no. 11, pp. 4839–4855, November 2006.

[9] D. Divsalar, H. Jin and R.J. McEliece, "Coding theorems for 'turbo-like' codes," *Proceedings of the 36th Allerton Conference on Communication, Control, and Computing*, pp. 201–210, Monticello, Illinois, September 23–25, 1998.

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

[11] 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.

[12] 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.

[13] P. M. Ebert, *Error Bounds for Parallel Communication Channels*, MIT, Ph.D. dissertation, August 1966. [Online]. Available: https://dspace.mit.edu/bitstream/1721.1/4295/RLE-TR-448-04743384.pdf.

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

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

[16] A. Guillen i Fabregas and G. Caire, "Coded modulation in the block-fading channel: coding theorems and code construction," *IEEE Trans. on Information Theory*, vol. 52, no. 1, pp. 91–114, January 2006.

[17] J. Ha, J. Kim and S. W. McLaughlin, "Rate-compatible puncturing of low-density parity-check codes," *IEEE Trans. on Information Theory*, vol. 50, no. 11, pp. 2824–2836, November 2004.

[18] C. H. Hsu and A. Anastasopoulos, "Asymptotic weight distributions of irregular repeat-accumulate codes," *Proceedings 2005 IEEE Global Telecommunications Conference*, vol. 3, no. 11, pp. 1147–1151, St. Louis, MO, USA, Nov. 2005.

[19] H. Jin and R. J. McEliece, "RA codes achieve AWGN channel capacity," *Lecture Notes in Computer Science*, vol. 1719, *Proceedings in Applied Algebra, Algebraic Algorithms and Error-Correcting Codes: 13th International Symposium (AAECC-13)*, pp. 10–18, Honolulu, Hawaii, USA, November 1999. M. Fossorier, H. Imai, S. Lin, and A. Poli (Eds.), published by Springer-Verlag Heidelberg.

[20] H. Jin and R. J. McEliece, "Irregular repeat-accumlate codes," *Proceedings Second International Conference on Turbo Codes and Related Topics*, pp. 1–8, Brest, France, September 2000.

[21] H. Jin and R. J. McEliece, "Coding theorems for turbo code ensembles," *IEEE Trans. on Information Theory*, vol. 48, no. 6, pp. 1451-1461, June 2002.

[22] A. Khandekar, *Graph-based Codes and Iterative Decoding*, Ph.D. dissertation, California Institute of Technology, Pasadena, CA, USA, June 2002. [Online]. Available:
`http://etd.caltech.edu/etd/available/etd-06202002-170522/unrestricted/thesis.pdf`.

[23] R. Liu, P. Spasojevic and E. Soljanin, "Reliable channel regions for good binary codes transmitted over parallel channels," *IEEE Trans. on Information Theory*, vol. 52, no. 4, pp. 1405–1424, April 2006.

[24] S. Litsyn and V. Shevelev, "Distance distributions in ensembles of irregular low-density parity-check codes," *IEEE Trans. on Information Theory*, vol. 49, no. 12, pp. 3140–3159, December 2003.

[25] R. J. McEliece, "How to compute weight enumerators for convolutional codes," in *Communications and Coding*, Tauton Research Studies, pp. 121–141, M. Darnel and B. Honary, Eds., Wiley, 1998.

[26] G. Miller and D. Burshtein, "Bounds on the maximum-likelihood decoding error probability of low-density parity-check codes," *IEEE Trans. on Information Theory*, vol. 47, no. 7, pp. 2696–2710, November 2001.

[27] L. C. Perez, J. Seghers and D. J. Costello, "A distance spectrum interpretation of turbo codes," *IEEE Trans. on Information Theory*, vol. 42, pp. 1698–1709, November 1996.

[28] H. D. Pfister, I. Sason and R. Urbanke, "Capacity-achieving ensembles for the binary erasure channel with bounded complexity," *IEEE Trans. on Information Theory*, vol. 51, no. 7, pp. 2352–2379, July 2005.

[29] H. D. Pfister and I. Sason, "Capacity-achieving ensembles of accumulate-repeat-accumulate codes for the erasure channel with bounded complexity," *IEEE Trans. on Information Theory*, vol. 53, no. 6, pp. 2088–2115, June 2007.

[30] 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.

[31] I. Sason and S. Shamai, "Gallager's 1963 bound: extensions and observations," Technical Report, CC No. 258, Technion, Israel, October 1998. [Online]. Available:
`http://www.ee.technion.ac.il/people/sason/CC258_{text, figures}.pdf`.

[32] 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.

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

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

[35] I. Sason and G. Wiechman, "On achievable rates and complexity of LDPC codes over parallel channels: Bounds and applications," *IEEE Trans. on Information Theory*, vol. 53, no. 2, pp. 580–598, February 2007.

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

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

[38] X. Wu, H. Xiang, and C. Ling, "New Gallager bounds in block-fading channels," *IEEE Trans. on Information Theory*, vol. 53, no. 2, pp. 684–694, February 2007.

[39] J. Zheng and S. L. Miller, "Performance analysis of coded OFDM systems over frequency-selective fading channels," *Proceedings 2003 IEEE Global Telecommunications Conference (GLOBECOM '03)*, pp. 1623–1627, San Francisco, CA, USA, December 1–5, 2003.