Gallager-Type Bounds for Parallel Channels with Applications to Modern Coding Techniques

Idan Goldenberg

Gallager-Type Bounds for Parallel Channels with Applications to Modern Coding Techniques

Research Thesis

Submitted in Partial Fulfillment of the Requirements

For the Degree of Master of Science

in Electrical Engineering

Idan Goldenberg

Submitted to the Senate of the Technion — Israel Institute of Technology

SHEBAT 5767 HAIFA FEBRUARY 2007

The Research Thesis Was Done Under The Supervision of Dr. Igal Sason in the Electrical Engineering Department

Acknowledgment

My sincere and wholehearted thanks go, first and foremost, to my supervisor, Dr. Igal Sason. Without his meticulous and consistent devotion to my guidance, this work would not be. He has also taught me a great deal about the fine art of preparing a lecture.

I would like to thank Prof. Shlomo Shamai and Dr. Ofer Amrani for serving on my examination committee.

Finally, I thank my wife, Tal Goldenberg, for her moral support and never-ending patience throughout this period. This thesis is devoted to her.

This work was supported by an annual grant from Intel Haifa.

To my wife, Tal.

Contents

List of Figures

Abstract

Some communication scenarios can be modelled as standard coded transmission over a set of parallel communication channels. These include transmission over block fading channels, rate-compatible puncturing of turbo-like codes, multi-carrier signaling and others. This thesis is focused on the performance analysis of binary linear block codes (or ensembles) whose transmission takes place over independent and memoryless parallel channels. New upper bounds on the maximum-likelihood (ML) decoding error probability are derived. These bounds include the generalization of the second version of the Duman and Salehi (DS2) bound to the case of parallel channels and a generalization of the classic 1961 Gallager bound to parallel channels. Optimized tilting measures for the new bound are derived. The connection between the generalized DS2 and the 1961 Gallager bounds, which was previously addressed for a single channel, is explored in the case of an arbitrary number of independent parallel channels. The generalization of the DS2 bound for parallel channels enables to re-derive specific bounds which were previously derived as special cases of the Gallager bound. The new bounds are applied to various ensembles of turbo-like codes, focusing especially on repeat-accumulate codes and their recent variations which possess low encoding and decoding complexity and exhibit remarkable performance under iterative decoding. 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. The improved bounds with their optimized tilting measures show, irrespectively of the block length of the codes, an improvement over the union bound and other previously reported bounds for independent parallel channels; this improvement is especially pronounced for moderate to large block lengths.

List of notation and abbreviations

ARA	:Accumulate-repeat-accumulate
AWGN	:Additive White Gaussian Noise
BER	:Bit error ratio
CSIR	:Channel side information at the receiver
i.i.d	:independent identically distributed
DS2	:Generalized second version of the Duman-Salehi Bound
IOWE	:Input-Output Weight Enumerator
IOWEF	:Input-Output Weight Enumeration Function
LDPC	:Low-Density Parity-Check
MBIOS	:Memoryless, Binary-Input and Output-Symmetric
ML	:Maximum Likelihood
RA	:Repeat and Accumulate
SF	:Shulman-Feder
SPARA	:Systematic and punctured Accumulate-repeat-accumulate
SPC	:Single parity-check
SPRA	:Systematic and punctured Repeat-accumulate

A_h	:Number of codewords of Hamming weight h
$A_{h_1,h_2,\ldots,h_J}$	:Split weight enumerator
$A_{w,h}$	:Input-output weight enumerator
α_j	:Probability of assigning a coded bit to channel j
C_j	:The capacity of channel j
γ	:Bhattacharyya constant
δ	:Normalized Hamming weight
$\frac{E_{\mathrm{b}}}{N_0}$	:Energy per bit to spectral noise density
$H(\cdot)$	:The binary entropy function to the natural base
J	:The number of parallel channels
k	:The dimension of a linear block code
n	:The length of a block code
N_0	:The one-sided spectral power density of the additive white Gaussian noise
$\Pr(A)$	:The probability of event A
P_{b}	:The bit error probability
P_{e}	:The block error probability
$\Psi_n^{(m)}(\underline{y})$	:Probability tilting measure
$Q(\cdot)$	:The Q-function
$r^{[\mathcal{C}]}(\delta)$	:Asymptotic exponent of the distance spectrum
R	:Code rate
R_0	:The cuttoff rate of a channel

Chapter 1

Introduction

Some modern communication systems are required to operate over multiple communication channels at once or over a single channel which varies with time. 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., [19, 43]), rate-compatible puncturing of turbo-like codes (see, e.g., [20, 39]), incremental redundancy retransmission schemes, cooperative coding, multi-carrier signaling (for performance bounds of coded orthogonal-frequency division multiplexing (OFDM) systems, see e.g., [44]), 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 minimum 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 [17]. Other examples of tight upper bounds include the generalized second version of the Duman-Salehi bound (often termed as the DS2 bound) [15, 40], the tangential sphere bound [33], the Shulman and Feder bound [41], and others. In this respect, it was shown by Sason and Shamai [40] that many reported upper bounds are special cases of the DS2 bound, including the 1961 Gallager bound [17]. For a comprehensive monograph on performance bounds of linear codes under ML decoding, the reader is referred to [37].

In his thesis [16], 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 block length. The exponents of the upper and lower bounds coincide for zero rate and for 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 [26] 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. [26] derived upper bounds on the ML decoding error probability which solely depend 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 individual transmission of these subsets over different channels. The analysis in [26] modifies the 1961 Gallager bound from [17, Chapter 3] and adapts this bounding technique for communication over parallel channels. However, the results presented in [26] 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 [41], simplified sphere bound [13], 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. Second, the DS2 bound ([14, 37, 40]) can be generalized to parallel channels.

Using the approach of the random mapper by Liu et al. [26], we derive a parallel-channel generalization of the DS2 bound [14, 37, 40] via two separate bounding techniques which yield two different bounds. 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 than the other. We re-examine, for the case of parallel channels, the well-known relations between this bound and the 1961 Gallager bound which exist for the single channel case [13, 40]. In this respect, it is shown 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.

The new bounds are used to obtain inner bounds on the boundary of the channel regions which are asymptotically (in the limit where we let the block length tend to infinity) attainable under ML decoding, and the results improve on those recently reported in [26]. 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 thesis is organized as follows. Chapter 2 deals with the calculation of the distance spectrum for some structured ensembles of turbo-like codes. The system model is presented in Chapter 3, as well as preliminary material related to our discussion. In Chapter 4, we generalize the DS2 bound for independent parallel channels using two different approaches. Chapter 5 presents the 1961 Gallager bound from [26], and considers its connection to the two versions of the DS2 bound, along with the optimization of its tilting measures. Chapter 6 presents some special cases of these upper bounds which are obtained as particular cases of the generalized bounds in Chapters 4 and 5. Attainable channel regions are derived in Chapter 7. Inner bounds on attainable channel regions for various ensembles of turbo-like codes and performance bounds for moderate block lengths are exemplified in Chapter 8. Finally, Chapter 9 concludes the thesis and considers topics for further research.

The results in this research work are also presented in [34], which was recently accepted for publication in the *IEEE Trans. on Information Theory* (as a full paper).

Chapter 2

Distance Properties of some Code Ensembles

2.1 Short overview

Bounds on the ML decoding error probability are often based on the distance properties of the considered codes or ensembles (see, e.g., [37] and references therein). The distance spectra and their asymptotic growth rates for various turbo-like ensembles have been studied in the literature, e.g., for ensembles of uniformly interleaved repeat-accumulate codes and variations [2, 11, 21], ensembles of uniformly interleaved turbo codes [4, 5, 28, 38], and ensembles of regular and irregular LDPC codes [8, 10, 17, 27]. In this Chapter, we present the distance properties of some turbo-like ensembles considered in this dissertation. We also consider as a reference the ensemble of fully random block codes which achieves capacity under ML decoding.

2.2 Preliminaries

Let us denote by $[\mathcal{C}(n)]$ an ensemble of codes of length n. We will also consider a *sequence of ensembles* $[\mathcal{C}(n_1)], [\mathcal{C}(n_2)], \ldots$, all of which 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

$$A_h^{[\mathcal{C}(n)]} \triangleq \frac{1}{|[\mathcal{C}(n)]|} \sum_{\mathcal{C} \in [\mathcal{C}(n)]} A_h^{\mathcal{C}} \ . \tag{2.1}$$

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

$$r^{\mathcal{C}}(\delta) \triangleq \frac{\ln A_h^{\mathcal{C}}}{n} \quad , \quad r^{[\mathcal{C}(n)]}(\delta) \triangleq \frac{\ln A_h^{[\mathcal{C}(n)]}}{n}. \tag{2.2}$$

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

$$r^{[\mathcal{C}]}(\delta) \triangleq \lim_{n \to \infty} r^{[\mathcal{C}(n)]}(\delta) \ . \tag{2.3}$$

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, and it is related to the distance spectrum by $A_h = \sum_{w=0}^{k} A_{w,h}$. Another quantity which we will be interested in is the *weighted distance spectrum* which is defined by

$$A'_h = \sum_{w=0}^{k} \frac{w}{k} A_{w,h}. \tag{2.4}$$

The weighted distance spectrum will be useful later for expressing bounds on the bit error probability, while the distance spectrum will used to express bounds on the block (decoding) error probability, both under ML decoding. Since both these quantities can be easily derived from the IOWE, we will focus on calculating the IOWE for the considered code ensembles. In this context, one considers the average IOWE over the code ensemble.

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

2.3 The ensemble of fully random binary block codes

Consider the ensemble of binary random codes $[\mathcal{RB}(n,R)]$, which consists of all binary codes of length n and rate R. In the case where $n \to \infty$, we use the notation $[\mathcal{RB}(R)]$ to express the asymptotic growth rate of the distance spectrum. For this ensemble, the following well-known equalities hold:

$$\begin{aligned} A_{w,h}^{[\mathcal{RB}(n,R)]} &= \binom{nR}{w}\binom{n}{h}2^{-n} \\ A_{h}^{[\mathcal{RB}(n,R)]} &= \binom{n}{h}2^{-n(1-R)} \\ r^{[\mathcal{RB}(n,R)]}(\delta) &= \frac{\ln\binom{n}{h}}{n} - (1-R)\ln 2 \\ r^{[\mathcal{RB}(R)]}(\delta) &= H(\delta) - (1-R)\ln 2 \end{aligned} \tag{2.5}$$

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

2.4 Turbo Codes

Turbo Codes were introduced in [7] and have been shown to exhibit astonishing performance. Their discovery has sparked an immense amount of research dealing with their properties, structure and performance. The encoder of a Turbo Code consists of two (and sometimes more) constituent systematic encoders joined together by an interleaver. In this section, we will consider an *ensemble* of Turbo Codes; the ensemble is defined in terms of the interleaver which is selected at random with a uniform distribution from the set of all possible permutations. This ensemble is thus termed as the ensemble of *uniformly interleaved Turbo Codes*, and is defined with respect to a specific constituent encoder. A schematic diagram of the Turbo encoder we will consider is shown in Fig. 2.1. Benedetto and Montorsi [4] have shown that it is possible to calculate the IOWE of the Turbo Code ensemble if we know the IOWE of each of the constituent encoders. The ensemble of Turbo Codes with recursive systematic convolutional encoders as constituent codes has been shown (see, e.g, [7], [12]) to yield excellent performance. Therefore, we will focus on this ensemble

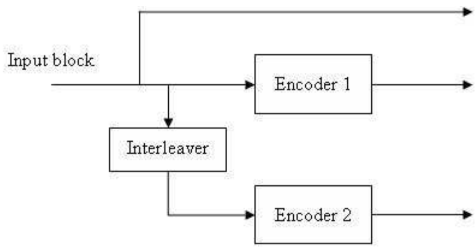


Figure 2.1: Schematic diagram of a Turbo encoder.

and show how to calculate its IOWE. The original Turbo Codes presented in [7] feature recursive systematic convolutional encoders as constituent codes. These codes combined with the iterative turbo decoding scheme provide results which are not far from theoretical limits. We will therefore analyze the performance of these codes using upper bounds on the ML decoding error probability. To this end, we calculate the IOWE of the ensemble of uniformly interleaved Turbo Codes with convolutional constituent codes.

2.4.1 The IOWE of a Convolutional Encoder

First, we turn our attention to the calculation of the IOWE of convolutional codes; this problem was solved by McElice [28] and we present the solution here.

We consider a sequence of length n at the output of a binary convolutional encoder which has m memory cells and therefore 2^m states. A *0-closed codepath of length n* is defined as a sequence of encoder states of length n in which the first and last states are the zero state. These codepaths are closely related to the non-zero codewords of the convolutional code. Thus, we will be interested in enumerating the 0-closed codepaths of length n; specifically, we will calculate the function

$$W(x, y, z) = \sum_{n \geq m} W_n(x, y) z^n, \tag{2.6}$$

called the *input-output weight enumerating function* (IOWEF), where

$$W_n(x, y) = \sum_{w,h \geq 0} A_{w,h} x^w y^h \tag{2.7}$$

and $A_{w,h}$ is the IOWE of the convolutional code which is terminated after a codeword of length n is created. To ensure proper termination of the encoder at the all-zero state, a codeword of length n is created by 'turning off' the input stream after $n-m$ input symbols and then forcing the encoder back into the all-zero state using a proper input sequence of length m. This is why the summation in (2.6) starts from $n=m$. As an example, consider the encoder of a rate 1/2 code appearing in Fig. 2.2(a). This encoder has $m=2$ memory cells. The encoder may also be described in terms of its *state diagram*, which depicts the outputs and state transitions as a function of the current state and input. The state diagram for the encoder in Fig. 2.2(a) is given in Fig. 2.2(b).

The analysis of input-output weight distributions is performed by enumerating all possible state sequences for codewords of the desired length n. This is done by breaking down the state sequence into a sequence of transitions and by analyzing the effect each transition has on the overall enumerator. The enumeration of a single state transition can be done using two enumeration matrices. The first is the *input-output incidence matrix* which is defined as a square matrix with number of rows equal to the number of encoder states, in which the (m,l)-th entry is a monomial of the form $x^i y^j$, where i and j are the input and output weights, respectively, associated with the transition from state m to state l, if this transition exists; if not, the (m,l)-th entry of the matrix is zero. For example, the state transition matrix associated with the encoder in Fig. 2.2(a) is given by

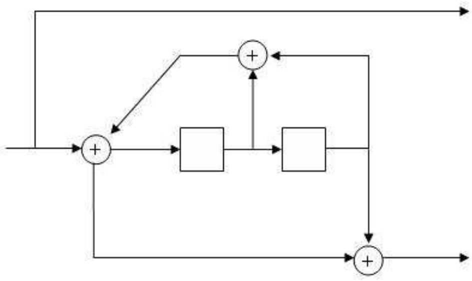
$$A(x,y) = \begin{array}{c c} & \begin{array}{cccc} 00 & 10 & 01 & 11 \end{array} \\ \begin{array}{c} 00 \\ 10 \\ 01 \\ 11 \end{array} & \left(\begin{array}{cccc} 1 & xy^2 & 0 & 0 \\ 0 & 0 & xy & y \\ xy^2 & 1 & 0 & 0 \\ 0 & 0 & y & xy \end{array} \right) \end{array} \tag{2.8}$$

The second enumeration matrix is the *output incidence matrix* $B(y)$ which is defined as

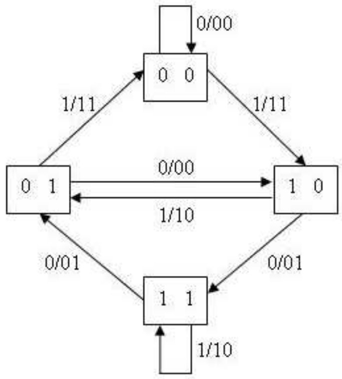
$$B(y) = A(1,y)$$

The IOWEF of the convolutional code is given by the following expression (see [28, Theorem 5.1])

$$W(x,y,z) = z^m \{(I - zA(x,y))^{-1} B(y)^m\}_{(0,0)} \tag{2.9}$$



(a) Encoder of a rate 1/2 convolutional code.



(b) State diagram of convolutional encoder.

Figure 2.2: (a) A convolutional encoder with transfer function $\left[1, \frac{1+D^2}{1+D+D^2}\right]$. (b) The corresponding state diagram. In this diagram "a/bc" on a state transition $i \rightarrow j$ means that given state i of the encoder and input bit a, bits b and c are the systematic and coded outputs, respectively, and the next state is j.

We illustrate this result by exemplifying the use of (2.9) on the encoder of Fig. 2.2(a). We have the state transition matrix $A(x, y)$ given in (2.8), and from it we calculate $B(y)$ according to (2.4.1). Using a symbolic manipulation program such as *Matlab*, it is easy to apply (2.9) to find that

$$W(x, y, z) = \frac{P(x, y, z)}{Q(x, y, z)} \qquad (2.10)$$

where

$$
\begin{aligned}
P(x,y,z) &= z^2 + (-xy + xy^5)z^3 + (-xy + x^2y^5 + xy^6 - x^2y^6)z^4 \\
&\quad +(-y^2 + x^2y^2 + xy^6 - x^3y^6)z^5
\end{aligned}
$$

and

$$
\begin{aligned}
Q(x,y,z) &= 1 + (-1 - xy)z + (xy - y^2 + x^2y^2 - x^3y^5)z^3 \\
&\quad +(y^2 - x^2y^2 - x^2y^6 + x^4y^6)z^4. \qquad (2.11)
\end{aligned}
$$

Using *Matlab*, the first few terms in the expansion of $W(x,y,z)$ as a power series in z can be found as

$$
\begin{aligned}
W(x,y,z) &= z^2 + (1 + xy^5)z^3 + (1 + (x + x^2)y^5 + xy^6)z^4 \\
&\quad +(1 + (x + x^2 + x^3)y^5 + (2x + x^2)y^6 + x^2y^7)z^5 \\
&\quad +\text{terms of order } z^6 \text{ and higher} \qquad (2.12)
\end{aligned}
$$

The remaining terms in the expansion of $W(x,y,z)$ can be obtained by noting that the form of the denominator in (2.11) implies that the individual enumerators $W_n(x,y)$ satisfy the fourth-order recursion

$$
\begin{aligned}
W_n &= (1 + xy)W_{n-1} + (-xy + y^2 - x^2y^2 + x^3y^5)W_{n-3} \qquad (2.13) \\
&\quad +(-y^2 + x^2y^2 + x^2y^6 - x^4y^6)W_{n-4} \quad \text{for } n \geq 6. \qquad (2.14)
\end{aligned}
$$

Therefore, we can use the recursion in (2.13) with the initial conditions

$$
\begin{aligned}
W_2(x,y) &= 1 \\
W_3(x,y) &= 1 + xy^5 \\
W_4(x,y) &= 1 + (x + x^2)y^5 + xy^6 \\
W_5(x,y) &= 1 + (x + x^2 + x^3)y^5 + (2x + x^2)y^6 + x^2y^7 \qquad (2.15)
\end{aligned}
$$

to obtain the value of W_n for any n by extending the recursion as far as desired.

2.4.2 The IOWE of a Turbo Code

A Turbo encoder which appears in Fig. 2.1 is the parallel concatenation of two constituent encoders. A *uniformly interleaved ensemble of Turbo codes* is the set of all possible Turbo codes with given constituent encoders, when considering all possible interleavers. In this setting, it is possible to compute the *average* IOWE of the Turbo code ensemble [4]. The average IOWE of an $(n_1 + n_2 - k, k)$ systematic Turbo code ensemble which is the parallel concatenation of an (n_1, k) systematic code (or ensemble) possessing an average IOWE $A^1_{w,h}$ with an (n_2, k) systematic code (or ensemble) possessing an average IOWE $A^2_{w,h}$ is given by

$$A_{w,h} = \frac{\displaystyle\sum_{h_1,h_2:\, h_1+h_2-w=h} A^1_{w,h_1} A^2_{w,h_2}}{\binom{k}{w}} \tag{2.16}$$

where we take into account that the systematic bits are transmitted only once. The asymptotic exponent of the distance spectrum for ensembles of convolutional and Turbo codes is given in [38].

2.5 Systematic and Non-Systematic Repeat-Accumulate Codes and Variations

In this section we will calculate the IOWE for three ensembles of turbo-like codes. These include the ensemble of repeat-accumulate (RA) codes and variations of this ensemble, one of which is an ensemble of accumulate-repeat-accumulate (ARA) codes. The encoders of these ensembles are shown in Fig. 2.3.

The component codes constructing these three ensembles are an accumulate code (i.e., a rate-1 differential encoder), a repetition code and a single parity-check (SPC) code. These components are serially concatenated in different combinations to create the encoders of these ensembles and hence, the IOWE of these codes can be expressed using the IOWE of their basic building blocks using the relations in [4, 5]. As a preparatory step, we introduce the IOWEs of the components.

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

$$A^{\mathrm{REP}(q)}_{w,d} = \binom{k}{w} \delta_{d,qw} \tag{2.17}$$

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

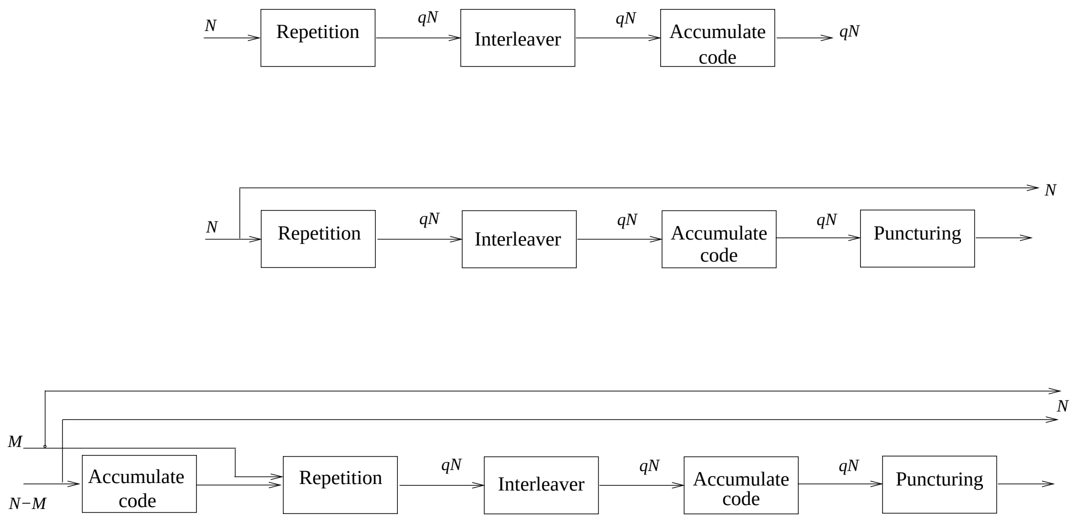


Figure 2.3: Systematic and Non-systematic RA and ARA codes. The interleavers of these ensembles are assumed to be chosen uniformly at random, and are of length qk where k designates the length of the input block (information bits) and q is the number of repetitions. The rates of all the ensembles is 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) where p is the puncturing period.

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

$$A_{w,d}^{\text{ACC}} = \binom{n-d}{\lfloor \frac{w}{2} \rfloor} \binom{d-1}{\lceil \frac{w}{2} \rceil - 1} \tag{2.18}$$

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 (2.18) 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)])

$$\begin{aligned} A(W,D) &= \sum_{w=0}^{np}\sum_{d=0}^{n} A_{w,d}^{\mathrm{SPC}(p)} W^w D^d \\ &= \Big[\mathrm{Even}\big((1+W)^p\big) + \mathrm{Odd}\big((1+W)^p\big)D\Big]^n \end{aligned} \tag{2.19}$$

where

$$\begin{aligned} \mathrm{Even}\big((1+W)^p\big) &= \frac{(1+W)^p + (1-W)^p}{2} \\ \mathrm{Odd}\big((1+W)^p\big) &= \frac{(1+W)^p - (1-W)^p}{2} \end{aligned} \tag{2.20}$$

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

To verify (2.19), 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).

Figure 2.4: Accumulate code with puncturing period $p = 3$ and an equivalent version of an SPC(p) code followed by an accumulate code.

The uniformly interleaved serial concatenation of an (N, k) code ensemble with average IOWE $A^1_{w,h}$ with an (n, N) code ensemble with average IOWE $A^2_{w,h}$ yields an (n, k) code ensemble with average IOWE given by [5]

$$A_{w,h} = \sum_{l=0}^{N} \frac{A^1_{w,l} A^2_{l,h}}{\binom{N}{l}}. \tag{2.21}$$

In what follows, we will capitalize on this relation for the calculation of the IOWEs of the three ensembles shown in Fig 2.3.

2.5.1 Non-systematic repeat-accumulate codes

The encoder of the ensemble of uniformly interleaved and non-systematic repeat-accumulate (NSRA) codes [11] is shown Fig. 2.3 (a). The ensemble [NSRA(k,q)] is defined as the set of all possible RA codes when considering the different permutations of the interleaver. As a side note, we can show that the number of codes in the ensemble is exactly $\frac{(qk)!}{(q!)^k k!}$. This can be seen by realizing that there is a total of $(qk)!$ ways to permute the order of qk bits. However, permuting the q repetitions of any of the k information bits does not affect the result of the interleaving, so there are $\frac{(qk)!}{(q!)^k}$ possible ways for the interleaving. Strictly speaking, by permuting the information bits, the vector space of the code does not change, which then yields that there are $\frac{(qk)!}{(q!)^k k!}$ distinct RA codes of dimension k and number of repetitions q.

The (average) IOWE of the ensemble of uniformly interleaved NSRA codes was originally derived in [11, Section 5]. This ensemble is simply the serial concatenation of a repetition code with an accumulator, so its IOWE can be obtained by using (2.17) and (2.18) in (2.21) which gives

$$A_{w,h}^{\mathrm{NSRA}(k,q)} = \frac{\binom{k}{w}\binom{qk-h}{\lfloor \frac{qw}{2} \rfloor}\binom{h-1}{\lceil \frac{qw}{2} \rceil - 1}}{\binom{qk}{qw}}. \tag{2.22}$$

The distance spectrum of the ensemble is therefore given by

$$A_h^{\mathrm{NSRA}(k,q)} = \sum_{w=1}^{\min(k, \lfloor \frac{2h}{q} \rfloor)} \frac{\binom{k}{w}\binom{qk-h}{\lfloor \frac{qw}{2} \rfloor}\binom{h-1}{\lceil \frac{qw}{2} \rceil - 1}}{\binom{qk}{qw}}, \quad \left\lceil \frac{q}{2} \right\rceil \leq h \leq qk - \left\lfloor \frac{q}{2} \right\rfloor$$

where $A_h^{\mathrm{NSRA}(k,q)} = 0$ for $1 \leq h < \lceil \frac{q}{2} \rceil$, and $A_0^{\mathrm{NSRA}(k,q)} = 1$ 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 [22])

$$\begin{aligned} r^{[\mathrm{NSRA}(q)]}(\delta) &\triangleq \lim_{k\to\infty} r^{[\mathrm{NSRA}(k,q)]}(\delta) \\ &= \max_{0 \leq u \leq \min(2\delta, 2-2\delta)} \left\{ -\left(1-\frac{1}{q}\right) H(u) + \right. \\ &\qquad \left. (1-\delta) H\left(\frac{u}{2(1-\delta)}\right) + \delta H\left(\frac{u}{2\delta}\right) \right\}. \end{aligned} \tag{2.23}$$

2.5.2 Systematic and Punctured Repeat-Accumulate Codes

The second ensemble we consider is the ensemble of systematic and punctured repeat-accumulate (SPRA) codes, where the systematic branch is added and puncturing is performed on the coded bits. The notation SPRA(k, p, q) will be used for this ensemble, and we will consider the case 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 $p = 3$). We rely on the concepts of the analysis introduced in [2] and on the serial concatenation formula (2.21) for the calculation of the average IOWE of the ensemble of uniformly interleaved SPRA codes; the asymptotic growth rate of the distance spectrum is also calculated.

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. 2.4 which was originally shown in [2, Fig. 2]). Hence, for the uniformly interleaved ensemble of SPRA(k,3,6) codes, we are interested in the IOWE of the SPC(3) code. For the case where $p = 3$, (2.20) gives

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

and (2.19) thus gives the following IOWE of the SPC(3) code [2, Eq. (15)]:

$$A_{w,d}^{\text{SPC}(3)} = \binom{n}{d} \sum_{j=0}^{n} \sum_{i=\max(0,j-n+d)}^{\min(j,d)} \binom{d}{i} \binom{n-d}{j-i} 3^{d+j-2i} \, \delta_{w,2j+d}. \tag{2.24}$$

We rely here on the equivalence shown in Fig. 2.4, 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. 2.4) are permuted by the uniform interleaver which is placed after the repetition code (see Fig. 2.3 (b)), then the average IOWE of this ensemble remains unaffected by placing an additional uniform interleaver between the SPC and the inner accumulate codes, which is of length $\frac{qk}{p} = 2k$. By placing the additional interleaver, 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 Eq. (2.21), i.e.,

$$A_{w,d}^{\text{ACC}(3)} = \sum_{h=0}^{2k} \frac{A_{w,h}^{\text{SPC}(3)} A_{h,d}^{\text{ACC}}}{\binom{2k}{h}}. \tag{2.25}$$

The substitution of (2.18) and (2.24) into (2.25) gives

$$A_{w,d}^{\text{ACC}(3)} = \sum_{h=0}^{2k} \sum_{j=0}^{2k} \sum_{i=\max(0,j-2k+h)}^{\min(j,h)} \Bigg\{ \binom{h}{i} \binom{2k-h}{j-i} \binom{2k-d}{\lfloor \frac{h}{2} \rfloor} \binom{d-1}{\lceil \frac{h}{2} \rceil - 1} 3^{h+j-2i} \, \delta_{w,2j+h} \Bigg\}. \tag{2.26}$$

Note that (2.26) is similar to [2, Eq. (19)], except that k in the latter equation is replaced by $2k$ in (2.26). 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] for a code of rate one-half.

Since there is a uniform interleaver of length qk 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

$$\begin{aligned} A_{w,d}^{\text{SPRA}(k,3,6)} &= \sum_{l=0}^{6k} \frac{A_{w,l}^{\text{REP}(6)} A_{l,d-w}^{\text{ACC}(3)}}{\binom{6k}{l}} \\ &= \frac{\binom{k}{w} A_{6w,d-w}^{\text{ACC}(3)}}{\binom{6k}{6w}} \end{aligned} \tag{2.27}$$

where the last equality is due to the equality in (2.17). Substituting (2.26) in the RHS of (2.27) gives the average IOWE of the ensemble as

$$A_{w,d}^{\text{SPRA}(k,3,6)} = \frac{\binom{k}{w}}{\binom{6k}{6w}} \sum_{h=0}^{2k} \sum_{j=0}^{2k} \sum_{i=\max(0,j-2k+h)}^{\min(j,h)} \Bigg\{ \binom{h}{i} \binom{2k-h}{j-i} \binom{2k-d+w}{\lfloor \frac{h}{2} \rfloor} \binom{d-w-1}{\lceil \frac{h}{2} \rceil - 1} 3^{h+j-2i} \, \delta_{6w,2j+h} \Bigg\} \tag{2.28}$$

Having obtained the IOWE of this ensemble, we turn to the calculation of the asymptotic growth rate of the distance spectrum. This is obtained by normalizing the logarithm of the average distance spectrum of the considered ensemble by $n = 3k$ and letting k tend to infinity.

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

$$A_d = \sum_{w=0}^{k} A_{w,d} \tag{2.29}$$

where the IOWE $A_{w,d}$ is given by (2.28). Note that unless

$$\frac{w}{k} = \frac{2j+h}{6k} = \frac{2\rho_2+\eta}{2} \tag{2.30}$$

the $A_{w,d}$ vanishes, and therefore it does not affect the sum in the RHS of (2.29). In the limit where $k \to \infty$, the asymptotic growth rate of the average distance spectrum for the uniformly interleaved ensemble of SPRA$(k,3,6)$ codes is obtained from (2.28) and (2.29) at it reads

$$\begin{aligned} r^{\text{SPRA}(3,6)}(\delta) &= \lim_{k\to\infty} \frac{1}{3k} \ln \sum_{w=0}^{k} A_{w,d} \\ &= \lim_{k\to\infty} \max_{h,i,j} \Bigg\{ \frac{1}{3k} \Bigg[kH\left(\frac{w}{k}\right) - 6kH\left(\frac{6w}{6k}\right) + hH\left(\frac{i}{h}\right) \\ &\qquad +(2k-h)H\left(\frac{j-i}{2k-h}\right) + (2k-d+w)H\left(\frac{h}{2(2k-d+w)}\right) \\ &\qquad +(d-w-1)H\left(\frac{\frac{h}{2}-1}{d-w-1}\right) + (h+j-2i)\ln 3 \Bigg] \Bigg\}. \end{aligned} \tag{2.31}$$

where we have used the well-known relation for the binomial coefficient

$$\lim_{n\to\infty} \frac{1}{n} \ln \binom{n}{\beta n} = H(\beta)\ , \quad 0 \le \beta \le 1. \tag{2.32}$$

Eq. (2.31) is more naturally expressed in terms of the normalized parameters

$$\delta \triangleq \frac{d}{3k}, \quad \eta \triangleq \frac{h}{3k}, \quad \rho_1 \triangleq \frac{i}{3k}, \quad \rho_2 \triangleq \frac{j}{3k} \tag{2.33}$$

which appear in the summations in (2.28). The normalization by $3k$ yields that the new parameters satisfy

$$0 \le \delta \le 1, \quad 0 \le \eta \le \frac{2}{3}, \quad 0 \le \rho_2 \le \frac{2}{3}. \tag{2.34}$$

From the partial sum w.r.t. the index i in the RHS of (2.28), dividing the terms in the inequality

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

by $3k$ gives

$$\max\left(0, \rho_2+\eta-\frac{2}{3}\right) \le \rho_1 \le \min(\rho_2, \eta). \tag{2.35}$$

Since the codes are systematic and the input block is k bits long, the terms which contribute to the IOWE in the RHS of (2.28) satisfy

$$w \leq \min(d, k), \quad 6w = 2j + h \tag{2.36}$$

and, from (2.33), multiplying (2.36) by $\frac{1}{3k}$ gives

$$\frac{2\rho_2 + \eta}{6} \leq \min\Big(\delta, \frac{1}{3}\Big). \tag{2.37}$$

From the binomial coefficients which appear in the RHS of (2.28), it is required that

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

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

$$\eta - \rho_2 + 3\delta \leq 2, \quad \rho_2 + 2\eta \leq 3\delta. \tag{2.38}$$

The asymptotic growth rate of the distance spectrum for the ensemble of uniformly interleaved SPRA$(k, 3, 6)$ codes is therefore given by

$$\begin{aligned} r^{\text{SPRA}(3,6)}(\delta) = \max_{\eta,\rho_1,\rho_2} \Bigg\{ & -\frac{5}{3}\, H\left(\frac{2\rho_2+\eta}{2}\right) + \eta\, H\left(\frac{\rho_1}{\eta}\right) + \left(\frac{2}{3} - \eta\right)\, H\left(\frac{\rho_2 - \rho_1}{\frac{2}{3} - \eta}\right) \\ & + \left(\frac{2}{3} - \delta + \frac{2\rho_2+\eta}{6}\right) H\left(\frac{\eta}{2\left(\frac{2}{3} - \delta + \frac{2\rho_2+\eta}{6}\right)}\right) \\ & + \left(\delta - \frac{2\rho_2+\eta}{6}\right) H\left(\frac{\eta}{2\left(\delta - \frac{2\rho_2+\eta}{6}\right)}\right) \\ & + (\eta + \rho_2 - 2\rho_1) \ln 3 \Bigg\} \end{aligned} \tag{2.39}$$

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

$$\begin{aligned} & 0 \leq \eta \leq \frac{2}{3}, \quad 0 \leq \rho_2 \leq \frac{2}{3}, \quad 2\rho_2 + \eta \leq 6\delta, \quad \rho_2 + 2\eta \leq 3\delta \\ & \max\Big(0, \rho_2 + \eta - \frac{2}{3}\Big) \leq \rho_1 \leq \min(\rho_2, \eta), \quad \eta - \rho_2 + 3\delta \leq 2. \end{aligned} \tag{2.40}$$

2.5.3 Systematic and Punctured Accumulate-Repeat-Accumulate Codes

The last ensemble we will consider is that of uniformly interleaved, systematic and punctured accumulate-repeat-accumulate (SPARA) codes which appears in Fig 2.3(c). This ensemble is similar to the SPRA ensemble except that the coded branch is preceded by an additional outer accumulator which appears in front of the repetition code. Only $k-M$ of the input bits are passed through this extra accumulator, and the remaining M are passed directly to the repetition code. We thus use the notation SPARA(k, M, p, q) for the ensemble. In order to calculate the IOWE for this ensemble, we first deal with the precoder. The precoder is a binary linear block code whose first $k-M$ input bits are accumulated and the other M input bits remain unchanged. The IOWE of this precoder, call it $\texttt{Pre}(k, M)$, is given by

$$\begin{aligned} A_{w,d}^{\text{Pre}(k,M)} &= \sum_{m=0}^{M} \binom{M}{m} A_{w-m,d-m}^{\text{ACC}} \\ &= \sum_{m=0}^{M} \left\{ \binom{M}{m} \binom{k-M-d+m}{\lfloor \frac{w-m}{2} \rfloor} \binom{d-m-1}{\lceil \frac{w-m}{2} \rceil - 1} \right\} \end{aligned} \tag{2.41}$$

where the last equality relies on (2.18). As we have mentioned before for the case of SPRA codes, an additional uniform interleaver placed between the precoder and the following stages of the SPARA encoder does not affect the average IOWE; this ensemble can therefore be viewed as the serially concatenated, with a uniform interleaver of length k, placed between the precoder and the repetition code in Fig. 2.3 (c) (in addition to the uniform interleaver which is placed after the repetition code). Moreover, referring to the SPRA ensemble whose components are REP(6) and ACC(3), the input bits (which are provided by the precoder to the second stage in Fig. 2.3 (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. 2.3 (c) is given by

$$A_{w,d}^{\text{SPARA}(k,M,3,6)} = \sum_{l=0}^{k} \frac{A_{w,l}^{\text{Pre}(k,M)} \; A_{l,d-w+l}^{\text{SPRA}(k,3,6)}}{\binom{k}{l}} \,. \tag{2.42}$$

By substituting (2.28) and (2.41) into (2.42), one obtains the average IOWE of the ensemble as

$$\begin{aligned} A_{w,d}^{\text{SPARA}(k,M,3,6)} &= \sum_{m=0}^{M}\sum_{l=0}^{k}\sum_{h=0}^{2k}\sum_{j=0}^{2k}\sum_{i=\max(0,j-2k+h)}^{\min(j,h)} \left\{ \frac{\binom{M}{m}\binom{k-M-l+m}{\lfloor\frac{w-m}{2}\rfloor}\binom{l-m-1}{\lceil\frac{w-m}{2}\rceil-1}}{\binom{6k}{6l}} \right. \\ &\quad \binom{h}{i}\binom{2k-h}{j-i}\binom{2k-d+w}{\lfloor\frac{h}{2}\rfloor} \\ &\quad \left. \binom{d-w-1}{\lceil\frac{h}{2}\rceil-1} 3^{h+j-2i}\ \delta_{6l,2j+h} \right\}. \end{aligned} \tag{2.43}$$

The asymptotic growth rate of the distance spectrum of this ensemble is obtained by the calculation of the limit

$$r^{\text{SPARA}(\alpha,3,6)}(\delta) = \lim_{k\to\infty}\frac{1}{3k}\sum_{w=0}^{k} A_{w,d}^{\text{SPARA}(k,3k\alpha,3,6)}, \qquad \delta = \frac{d}{3k}\ \ (0 \le \delta \le 1)$$

where $\alpha \triangleq \frac{M}{3k}$ is a normalized parameter designating the fraction of input bits which do not pass through the outer accumulator. As in the case of SPRA codes, we will use the parameters ρ_1, ρ_2, δ, and η defined in (2.33) and also use the three additional parameters

$$\alpha \triangleq \frac{M}{3k}, \quad \varepsilon_1 \triangleq \frac{m}{3k}, \quad \varepsilon_2 \triangleq \frac{w-m}{3k}. \tag{2.44}$$

Since $M \le k$ we have that $0 \le \alpha \le \frac{1}{3}$. Now, $0 \le m \le M$, so we have the additional limitation $0 \le \varepsilon_1 \le \alpha$. Finally, the input weight w satisfies $0 \le w \le \min(d,k)$ so that we have another constraint which reads $0 \le \varepsilon_1 + \varepsilon_2 \le \min\Big(\delta, \frac{1}{3}\Big)$. After straightforward and tedious algebra which is similar to the calculations in Section 2.5.2, one obtains the following expression for the asymptotic growth rate of the average distance spectrum of the considered ensemble of uniformly interleaved SPARA codes:

$$
\begin{aligned}
r^{\mathrm{SPARA}(\alpha,3,6)}(\delta) \;=\; & \max_{\eta,\rho_1,\rho_2,\varepsilon_1,\varepsilon_2} \Bigg\{ \alpha H\left(\frac{\varepsilon_1}{\alpha}\right) + \left(\frac{2\rho_2+\eta}{6} - \varepsilon_1\right) H\left(\frac{\varepsilon_2}{2\left(\frac{2\rho_2+\eta}{6} - \varepsilon_1\right)}\right) \\
& + \left(\frac{1}{3} - \alpha - \frac{2\rho_2+\eta}{6} + \varepsilon_1\right) H\left(\frac{\varepsilon_2}{2\left(\frac{1}{3} - \alpha - \frac{2\rho_2+\eta}{6} + \varepsilon_1\right)}\right) \\
& + \eta\, H\left(\frac{\rho_1}{\eta}\right) - 2H\left(\frac{2\rho_2+\eta}{2}\right) + \left(\frac{2}{3} - \eta\right) H\left(\frac{\rho_2 - \rho_1}{\frac{2}{3} - \eta}\right) \\
& + \left(\frac{2}{3} - \delta + \varepsilon_1 + \varepsilon_2\right) H\left(\frac{\eta}{2\left(\frac{2}{3} - \delta + \varepsilon_1 + \varepsilon_2\right)}\right) \\
& + (\delta - \varepsilon_1 - \varepsilon_2) H\left(\frac{\eta}{2(\delta - \varepsilon_1 - \varepsilon_2)}\right) + (\eta + \rho_2 - 2\rho_1)\ln 3 \Bigg\} \qquad (2.45)
\end{aligned}
$$

where the five-parameter maximization is performed over the finite domain which is characterized by the following inequalities which stem from the limitations on the summation in (2.43):

$$
\begin{aligned}
& 0 \le \eta \le \frac{2}{3}, \quad 0 \le \rho_2 \le \frac{2}{3}, \quad 0 \le \varepsilon_1 \le \alpha, \\
& 0 \le \varepsilon_1 + \varepsilon_2 \le \min\left(\delta, \frac{1}{3}\right), \\
& \max\left(0, \rho_2 + \eta - \frac{2}{3}\right) \le \rho_1 \le \min(\rho_2, \eta), \\
& 0 \le \eta \le \min\left(\frac{4}{3} - 2\delta + 2(\varepsilon_1 + \varepsilon_2), 2\delta - 2(\varepsilon_1 + \varepsilon_2)\right), \\
& \varepsilon_2 \le \min\left(\frac{2}{3} - 2\alpha - \frac{2\rho_2+\eta}{3} + 2\varepsilon_1, \frac{2\rho_2+\eta}{3} - 2\varepsilon_1\right). \qquad (2.46)
\end{aligned}
$$

We exemplify some of the results appearing in this section in Fig. 2.5 which shows the asymptotic growth rates of the distance spectra for ensembles of repeat-accumulate codes considered in this chapter.

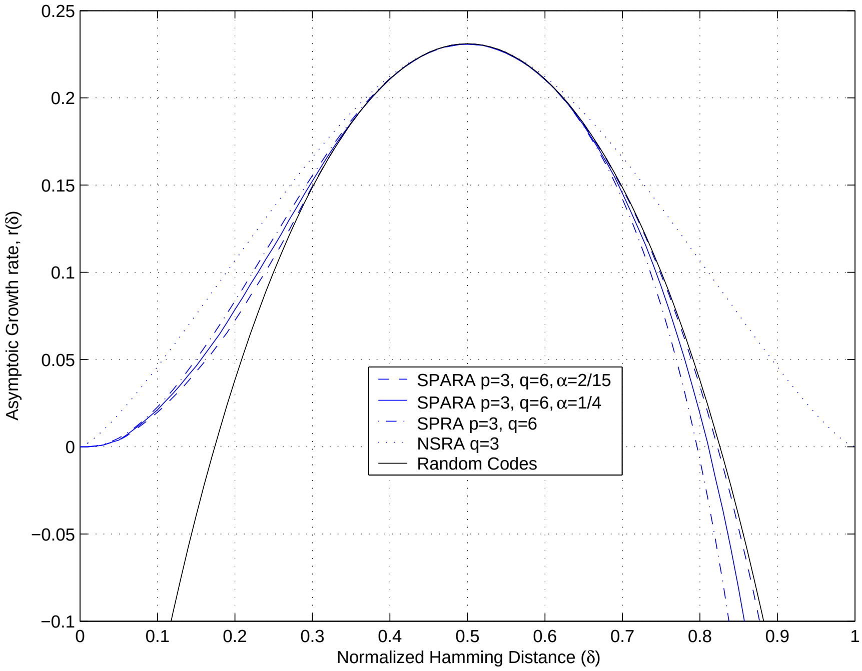


Figure 2.5: Comparison of asymptotic growth rates of the average distance spectra of ensembles of RA codes and variations.

Chapter 3

Definitions and Preliminaries

3.1 Short overview

In this chapter, 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 parallel channels.

3.2 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. 3.1.

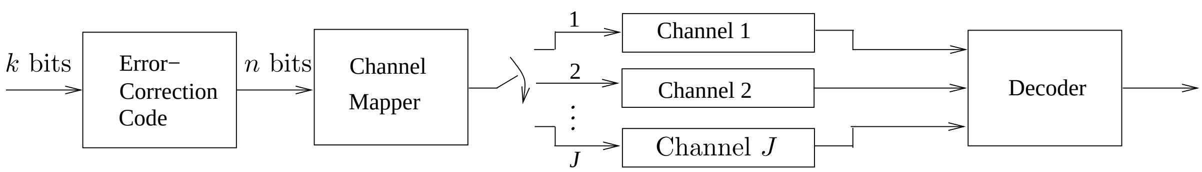


Figure 3.1: System model of parallel channels. A random mapper is assumed where every bit is assigned to one of the J channels; a bit is assigned to the j^{th} channel independently of the other bits and with probability α_j (where $\sum_{j=1}^{J} \alpha_j = 1$).

Using an error-correcting linear 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 has 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 when the specific channel mapper is known at the receiver. 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 chapter 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 [26] 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 in [26] the derivation of an upper bound for the parallel channels which is averaged over all possible channel assignments, and the bound can be 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.

3.3 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.

3.3.1 Cutoff Rate

The cutoff rate of an MBIOS channel is given by

$$R_0 = 1 - \log_2(1+\gamma) \tag{3.1}$$

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

$$\gamma \triangleq \sum_y \sqrt{p(y|0)p(y|1)}. \tag{3.2}$$

Clearly, for continuous-output channels, the sum in the RHS of (3.2) 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.,

$$\gamma = \sum_{J=1}^{J} \left\{ \alpha_j \sum_y \sqrt{p(y|0;j)p(y|1;j)} \right\}. \tag{3.3}$$

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

$$\begin{aligned} p(y|0;j) &= \frac{1}{\sqrt{2\pi}}\, e^{-\frac{(y+\sqrt{2\nu_j})^2}{2}} \\ p(y|1;j) &= \frac{1}{\sqrt{2\pi}}\, e^{-\frac{(y-\sqrt{2\nu_j})^2}{2}} \\ &-\infty < y < \infty,\ j = 1, \ldots, J \end{aligned} \tag{3.4}$$

where

$$\nu_j \triangleq R \left(\frac{E_\mathrm{b}}{N_0} \right)_j \tag{3.5}$$

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

$$\gamma = \sum_{j=1}^{J} \alpha_j e^{-\nu_j} \tag{3.6}$$

where ν_j is introduced in (3.5). From (3.1) and (3.6), the cutoff rate of J parallel binary-input AWGN channels is given by

$$R_0 = 1 - \log_2 \left(1 + \sum_{j=1}^{J} \alpha_j e^{-R\left(\frac{E_\mathrm{b}}{N_0}\right)_j} \right) \text{ bits per channel use.} \tag{3.7}$$

Consider the case of $J = 2$ parallel binary-input AWGN channels. Given the value of $\left(\frac{E_\mathrm{b}}{N_0}\right)_1$, and the code rate R (in bits per channel use), it is possible to calculate the value of $\left(\frac{E_\mathrm{b}}{N_0}\right)_2$ of the second channel which corresponds to the cutoff rate. To this end, we set R_0 in the LHS of (3.7) to R. Solving this equation gives

$$\left(\frac{E_\mathrm{b}}{N_0} \right)_2 = -\frac{1}{R} \ln \left(\frac{2^{1-R} - 1 - \alpha_1 e^{-R\left(\frac{E_\mathrm{b}}{N_0}\right)_1}}{\alpha_2} \right). \tag{3.8}$$

3.3.2 Capacity Limit

Let C_j designate the capacity (in bits per channel use) of the j-th MBIOS channel the set of J parallel MBIOS channels. 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

$$C = \sum_{j=1}^{J} \alpha_j C_j. \tag{3.9}$$

For the case of J parallel binary-input AWGN channels

$$C_j = 1 - \frac{1}{\sqrt{2\pi}\ln(2)} \int_{-\infty}^{\infty} e^{-\frac{(y-\beta_j)^2}{2}} \ln\left(1 + e^{-2\beta_j y}\right) \mathrm{d}y \quad \text{bits per channel use} \tag{3.10}$$

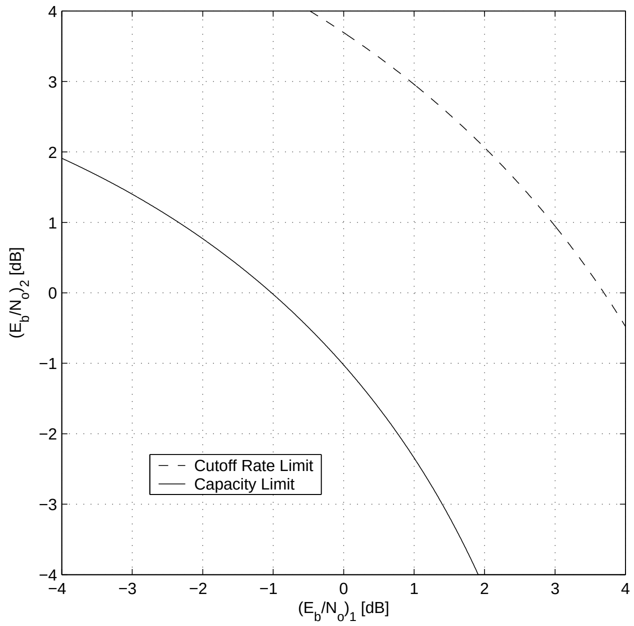


Figure 3.2: Attainable channel regions for two parallel binary-input AWGN channels, as determined by the cutoff rate and the capacity limit, referring to a code rate of one-third bits per channel use. It is assumed that each bit is randomly and independently assigned to one of these channels with equal probability (i.e., $\alpha_1 = \alpha_2 = \frac{1}{2}$).

where $\beta_j \triangleq \sqrt{2\nu_j}$ and ν_j is introduced in (3.5).

In order to simplify the numerical computation of the capacity, one can express each integral in (3.10) 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:

$$C_j = 1 - \frac{1}{\ln(2)} \left[\frac{2\beta e^{-\frac{\beta_j^2}{2}}}{\sqrt{2\pi}} - (2\beta_j^2 - 1)Q(\beta_j) + \sum_{k=0}^{\infty} \frac{(-1)^k \cdot \Delta^k a_0(j)}{2^{k+1}} \right], \quad j = 1, \ldots, J \tag{3.11}$$

where

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

and

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

(note that $\mathrm{erfcx}(x) \approx \frac{1}{\sqrt{\pi}} \cdot \frac{1}{x}$ for large values of x). The infinite sum in (3.11) 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 $\left(\frac{E_\mathrm{b}}{N_0}\right)_1$, and the code rate R (in bits per channel use), (3.9) and (3.10) enable one to calculate the value of $\left(\frac{E_\mathrm{b}}{N_0}\right)_2$ for the second channel, referring to the capacity limitation. To this end, one needs to set C in the LHS of (3.9) to the code rate R, and find the resulting value of $\left(\frac{E_\mathrm{b}}{N_0}\right)_2$ which corresponds to the capacity limit. The boundary of the capacity region is represented by the continuous curve in Fig. 3.2 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. (3.8)).

3.4 The DS2 Bound for a Single MBIOS Channel

The bounding technique of Duman and Salehi [14, 15] originates from the 1965 Gallager bound [18] which states that the conditional ML decoding error probability $P_{\mathrm{e}|m}$ given that a codeword $\underline{x}^m$ (of block length n) is transmitted is upper-bounded by

$$P_{\mathrm{e}|m} \leq \sum_{\underline{y}} p_n\left(\underline{y}|\underline{x}^m\right) \left(\sum_{m' \neq m} \left(\frac{p_n(\underline{y}|\underline{x}^{m'})}{p_n(\underline{y}|\underline{x}^m)} \right)^{\lambda} \right)^{\rho} \quad \lambda, \rho \geq 0 \qquad (3.12)$$

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 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

[40], 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 (3.12) can be rewritten in the following equivalent form:

$$P_{\mathrm{e}|m} \leq \sum_{\underline{y}} \Psi_n^{(m)}(\underline{y}) \left(\Psi_n^{(m)}(\underline{y})^{-\frac{1}{\rho}} p_n\left(\underline{y}|\underline{x}^m\right)^{\frac{1}{\rho}} \sum_{m' \neq m} \left(\frac{p_n(\underline{y}|\underline{x}^{m'})}{p_n(\underline{y}|\underline{x}^m)} \right)^{\lambda} \right)^{\rho} \qquad \lambda, \rho \geq 0. \tag{3.13}$$

Recalling that $\Psi_n^{(m)}$ is a probability measure, we invoke Jensen's inequality in (3.13) which gives

$$P_{\mathrm{e}|m} \leq \left(\sum_{m' \neq m} \sum_{\underline{y}} \Psi_n^{(m)}(\underline{y})^{1-\frac{1}{\rho}} p_n(\underline{y}|\underline{x}^m)^{\frac{1}{\rho}} \left(\frac{p_n(\underline{y}|\underline{x}^{m'})}{p_n(\underline{y}|\underline{x}^m)} \right)^{\lambda} \right)^{\rho}, \quad \begin{array}{c} 0 \leq \rho \leq 1 \\ \lambda \geq 0 \end{array} \tag{3.14}$$

which is the DS2 bound. This expression can be simplified (see, e.g., [40]) 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 (3.14) 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 (3.14) gives

$$\begin{aligned} P_{\mathrm{e}|0}(h) \quad \leq \quad (A_h)^{\rho} & \left\{ \left(\sum_y \psi(y)^{1-\frac{1}{\rho}} p(y|0)^{\frac{1}{\rho}} \right)^{n-h} \right. \\ & \left. \left(\sum_y \psi(y)^{1-\frac{1}{\rho}} p(y|0)^{\frac{1-\lambda\rho}{\rho}} p(y|1)^{\lambda} \right)^{h} \right\}^{\rho} \qquad \begin{array}{c} 0 \leq \rho \leq 1 \\ \lambda \geq 0 \end{array} . \end{aligned} \tag{3.15}$$

Clearly, for an MBIOS channel with continuous output, the sums in (3.15) 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 shown later even for the case of 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:

$$P_{\mathrm{e}} \leq \sum_{h=d_{\min}}^{n(1-R)} P_{\mathrm{e}|0}(h). \tag{3.16}$$

For the bit error probability, the same analysis applies except that the distance spectrum of the code is replaced by A'_h given in (2.4). This is due to the following lemma, derived by Divsalar [13].

Lemma 1 [13, Section III.C] Let $\mathcal{C}$ be a binary linear (n,k) block code transmitted over an MBIOS channel. Let $\mathcal{C}(w)$ designate a sub-code of $\mathcal{C}$ containing the all-zero codeword plus all the codewords encoded by an information block of Hamming weight w. Then, the conditional bit error probability of $\mathcal{C}$ under ML decoding, given that the all-zero codeword is transmitted, is upper bounded by

$$P_{\mathrm{b}|0} \leq \sum_{\underline{y}} p_n(\underline{y}|\underline{0}) \left(\sum_{w=1}^{k} \left(\frac{w}{k}\right) \sum_{\substack{\underline{c}\in\mathcal{C}(w) \\ \underline{c}\neq\underline{0}}} \left(\frac{p_n(\underline{y}|\underline{c})}{p_n(\underline{y}|\underline{0})} \right)^{\lambda} \right)^{\rho} \qquad \begin{array}{c} \lambda \geq 0 \\ 0 \leq \rho \leq 1 \end{array} . \tag{3.17}$$

The following proof of the lemma was brought in [42] and is a simplified version of the proof in [13].

Proof. The conditional bit error probability under ML decoding can be expressed as

$$P_{\mathrm{b}|0} = \sum_{\underline{y}} \left(\frac{w_0(\underline{y})}{k} \right) p_n(\underline{y}|\underline{0}) \tag{3.18}$$

where $w_0(\underline{y})$ designates the weight of the information bits in the decoded codeword, given that the all-zero codeword is transmitted and the received vector is $\underline{y}$. In particular, if the received vector $\underline{y}$ is within the decision region of the all-zero codeword, then $w_0(\underline{y}) = 0$. Now, we have the following inequalities.

$$\begin{aligned}
\frac{w_0(\underline{y})}{k} &\leq \left(\frac{w_0(\underline{y})}{k} \right)^{\rho}, \qquad 0 \leq \rho \leq 1 \\
&\stackrel{\text{(a)}}{\leq} \left\{ \left(\frac{w_0(\underline{y})}{k} \right) \sum_{\substack{\underline{c} \in \mathcal{C}(w_0(\underline{y})) \\ \underline{c} \neq \underline{0}}} \left(\frac{p_n(\underline{y}|\underline{c})}{p_n(\underline{y}|\underline{0})} \right)^{\lambda} \right\}^{\rho} \qquad \lambda \geq 0 \\
&\leq \left\{ \sum_{w=1}^{k} \left(\frac{w}{k} \right) \sum_{\substack{\underline{c} \in \mathcal{C}(w) \\ \underline{c} \neq \underline{0}}} \left(\frac{p_n(\underline{y}|\underline{c})}{p_n(\underline{y}|\underline{0})} \right)^{\lambda} \right\}^{\rho}.
\end{aligned} \tag{3.19}$$

Inequality (a) holds since the received vector $\underline{y}$ must fall in the decision region of some codeword $\underline{c}$ which is encoded by information bits of total Hamming weight $w_0(\underline{y})$; hence, the quotient $\left(\frac{p_n(\underline{y}|\underline{c})}{p_n(\underline{y}|\underline{0})} \right)$ is larger than 1 while the other terms in the sum are simply non-negative. The third inequality holds because of adding more non-negative terms to the sum. The lemma follows by substituting (3.19) into the RHS of (3.18). $\square$

Inequality (3.17) can be thought of as the counterpart to inequality (3.12), for the case where $m = 0$, where the former inequality relates to the bit error probability and the latter refers to the block error probability. With (3.17) as the starting point, the derivation of the DS2 bound on the block error probability appearing in the beginning of this section may be repeated in order to get a bound on the bit error probability. The result (see, [36, 37]) is that the conditional DS2 bound on the bit error probability is identical to the DS2 bound on the block error probability, except that the distance spectrum of the code A_h appearing in the RHS of (3.15) is replaced by A'_h given in (2.4). 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 (3.15) which yields that $\mathbb{E}[(A_h)^\rho] \leq (\mathbb{E}[A_h])^\rho$ for $0 \leq \rho \leq 1$.

The application of the DS2 bound to a single MBIOS channel is discussed in further details in [14, 36, 40] and the tutorial paper [37, Chapter 4].

Chapter 4

Generalized DS2 Bounds for Parallel Channels

4.1 Short overview

In this chapter, 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 forms. 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.

4.2 Generalizing the DS2 bound to Parallel Channels: First Approach

4.2.1 Derivation of the bound

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

$$p_n\left(\underline{y}|\underline{x}^m\right) = \prod_{j=1}^{J} \prod_{i \in \mathcal{I}(j)} p(y_i|x_i^{(m)}; j) \tag{4.1}$$

which we can plug into (3.14) 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 (3.14) so that it can be expressed as a J-fold product of one-dimensional probability tilting measures, i.e.,

$$\Psi_n^{(m)}(\underline{y}) = \prod_{j=1}^{J} \prod_{i \in \mathcal{I}(j)} \psi^{(m)}(y_i; j). \tag{4.2}$$

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 [26], 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 (4.1) and

(4.2) in (3.14), we obtain

$$
\begin{aligned}
P_{\mathrm{e}} &= P_{\mathrm{e}|0} \\
&\le \left\{ \sum_{h_1=0}^{|\mathcal{I}(1)|} \cdots \sum_{h_J=0}^{|\mathcal{I}(J)|} \sum_{\underline{y}} A_{h_1,h_2,\ldots,h_J} \prod_{j=1}^{J} \prod_{i \in \mathcal{I}(j)} \psi(y_i;j)^{1-\frac{1}{\rho}} p(y_i|0;j)^{\frac{1}{\rho}} \left(\frac{p(y_i|x_i;j)}{p(y_i|0;j)} \right)^{\lambda} \right\}^{\rho} \\
&= \left\{ \sum_{h_1=0}^{|\mathcal{I}(1)|} \cdots \sum_{h_J=0}^{|\mathcal{I}(J)|} A_{h_1,h_2,\ldots,h_J} \prod_{j=1}^{J} \prod_{i \in \mathcal{I}(j)} \sum_{y_i} \psi(y_i;j)^{1-\frac{1}{\rho}} p(y_i|0;j)^{\frac{1}{\rho}} \left(\frac{p(y_i|x_i;j)}{p(y_i|0;j)} \right)^{\lambda} \right\}^{\rho} \\
&= \left\{ \sum_{h_1=0}^{|\mathcal{I}(1)|} \cdots \sum_{h_J=0}^{|\mathcal{I}(J)|} A_{h_1,h_2,\ldots,h_J} \prod_{j=1}^{J} \left(\sum_{y} \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1-\lambda\rho}{\rho}} p(y|1;j)^{\lambda} \right)^{h_j} \right. \\
&\qquad \left. \prod_{j=1}^{J} \left(\sum_{y} \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1}{\rho}} \right)^{|\mathcal{I}(j)|-h_j} \right\}^{\rho}, \quad \begin{array}{c} 0 \le \rho \le 1 \\ \lambda \ge 0 \end{array}. \qquad (4.3)
\end{aligned}
$$

We note that the bound in (4.3) 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 (4.3) with respect to all possible channel assignments, we get the following bound on the average ML decoding error probability:

$$
\begin{aligned}
P_{\mathrm{e}} &\le \mathbf{E} \left\{ \sum_{h_1=0}^{n_1} \cdots \sum_{h_J=0}^{n_J} A_{h_1,h_2,\ldots,h_J} \prod_{j=1}^{J} \left(\sum_{y} \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1-\lambda\rho}{\rho}} p(y|1;j)^{\lambda} \right)^{h_j} \right. \\
&\qquad \left. \prod_{j=1}^{J} \left(\sum_{y} \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1}{\rho}} \right)^{n_j-h_j} \right\}^{\rho} \\
&= \sum_{\substack{n_j \ge 0 \\ \sum_j n_j = n}} \left\{ \sum_{h_1=0}^{n_1} \cdots \sum_{h_J=0}^{n_J} A_{h_1,h_2,\ldots,h_J} \prod_{j=1}^{J} \left(\sum_{y} \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1-\lambda\rho}{\rho}} p(y|1;j)^{\lambda} \right)^{h_j} \right. \\
&\qquad \left. \prod_{j=1}^{J} \left(\sum_{y} \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1}{\rho}} \right)^{n_j-h_j} \right\}^{\rho} P_{\underline{N}}(\underline{n}) \qquad (4.4)
\end{aligned}
$$

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 (4.4) and changing the order of summation give

$$
\begin{aligned}
P_{\mathrm{e}} \;\le\; & \Bigg\{ \sum_{\substack{n_j \ge 0 \\ \sum n_j = n}} \sum_{h=0}^{n} \sum_{\substack{h_1 \le n_1, \ldots, h_J \le n_J \\ h_1 + \ldots + h_J = h}} A_{h_1, h_2, \ldots, h_J} P_{\underline{N}}(\underline{n}) \\
& \qquad \prod_{j=1}^{J} \left(\sum_y \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1-\lambda\rho}{\rho}} p(y|1;j)^{\lambda} \right)^{h_j} \\
& \qquad \prod_{j=1}^{J} \left(\sum_y \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1}{\rho}} \right)^{n_j - h_j} \Bigg\}^{\rho}, \quad \begin{array}{c} 0 \le \rho \le 1 \\ \lambda \ge 0 \end{array}
\end{aligned} \tag{4.5}
$$

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 $0 \le h_j \le n_j$). 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

$$
\begin{aligned}
& P_{\underline{N}}(\underline{n}) = \binom{n}{n_1, n_2, \ldots, n_J} \alpha_1^{n_1} \alpha_2^{n_2} \ldots \alpha_J^{n_J} \\
& P_{\underline{H}|\underline{N}}(\underline{h}|\underline{n}) = \frac{\binom{h}{h_1, \ldots, h_J} \binom{n-h}{n_1 - h_1, \ldots, n_J - h_J}}{\binom{n}{n_1, \ldots, n_J}} \\
& A_{h_1, h_2, \ldots, h_J} P_{\underline{N}}(\underline{n}) \\
& = A_h \, P_{\underline{H}|\underline{N}}(\underline{h}|\underline{n}) \, P_{\underline{N}}(\underline{n}) \\
& = A_h \, \alpha_1^{n_1} \alpha_2^{n_2} \ldots \alpha_J^{n_J} \binom{h}{h_1, \ldots, h_J} \binom{n-h}{n_1 - h_1, \ldots, n_J - h_J}
\end{aligned} \tag{4.6}
$$

and the substitution of (4.6) in (4.5) gives

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

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

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

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

$$
\begin{aligned}
P_{\mathrm{e}} \;\le\; & \Bigg\{ \sum_{h=0}^{n} A_h \left(\sum_{j=1}^{J} \alpha_j \sum_y \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1-\lambda\rho}{\rho}} p(y|1;j)^{\lambda} \right)^{h} \\
& \qquad \left(\sum_{j=1}^{J} \alpha_j \sum_y \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1}{\rho}} \right)^{n-h} \Bigg\}^{\rho} \qquad \begin{array}{c} 0 \le \rho \le 1 \\ \lambda \ge 0 \\ \sum_y \psi(y;j) = 1 \\ j = 1 \dots J \end{array}
\end{aligned} \tag{4.7}
$$

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 $\overline{P_e}$ is obtained by replacing A_h in (4.7) with the average distance spectrum $\overline{A_h}$ (this can be verified by noting that the function $f(t) = t^\rho$ is convex for $0 \le \rho \le 1$ and by invoking Jensen's inequality in (4.7)).

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 [37, 40]. 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

$$\Psi_n^{(m)}(\underline{y}) = \frac{G_n^{(m)}(\underline{y}) p_n(\underline{y}|\underline{x}^m)}{\sum_{\underline{y}'} G_n^{(m)}(\underline{y}') p_n(\underline{y}'|\underline{x}^m)} . \tag{4.8}$$

Substituting (4.8) in (3.14) gives

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

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 (4.3)-(4.7) and averaging as before also over all the codebooks of the ensemble, we obtain the following upper bound on the ML decoding error probability:

$$
\begin{aligned}
P_{\mathrm{e}} = P_{\mathrm{e}|0} \quad \leq \quad & \left\{ \sum_{h=0}^{n} A_h \left[\sum_{j=1}^{J} \alpha_j \left(\sum_y g(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{1-\lambda} p(y|1;j)^{\lambda} \right) \right. \right. \\
& \left. \left(\sum_y g(y;j) p(y|0;j) \right)^{\frac{1-\rho}{\rho}} \right]^{h} \left[\sum_{j=1}^{J} \alpha_j \left(\sum_y g(y;j)^{1-\frac{1}{\rho}} p(y|0;j) \right) \right. \\
& \left. \left. \left(\sum_y g(y;j) p(y|0;j) \right)^{\frac{1-\rho}{\rho}} \right]^{n-h} \right\}^{\rho}, \quad \begin{array}{c} 0 \leq \rho \leq 1 \\ \lambda \geq 0 \end{array}. \qquad (4.9)
\end{aligned}
$$

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 *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.

4.2.2 Optimization of the Tilting Measures

In the following, we find optimized tilting measures $\{\psi(\cdot;j)\}_{j=1}^{J}$ which minimize the DS2 bound (4.7). The following calculation is a possible generalization of the analysis in [40] 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 [26, 40], 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

$$P_{\rm e} \leq \sum_{h=0}^{n} P_{{\rm e}|0}(h). \tag{4.10}$$

As the code $\mathcal{C}$ is linear, $P_{{\rm e}|0}(h) = 0$ 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 (4.7) gives the following upper bound on the conditional error probability of the subcode $\mathcal{C}_h$:

$$\begin{aligned} P_{{\rm e}|0}(h) \;\; &\leq \;\; (A_h)^{\rho} \left\{ \left(\sum_{j=1}^{J} \alpha_j \sum_{y} \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1-\lambda\rho}{\rho}} p(y|1;j)^{\lambda} \right)^{\delta} \right. \\ & \left. \left(\sum_{j=1}^{J} \alpha_j \sum_{y} \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1}{\rho}} \right)^{1-\delta} \right\}^{n\rho}, \quad \begin{array}{c} 0 \leq \rho \leq 1 \\ \lambda \geq 0 \end{array} \quad \delta \triangleq \frac{h}{n}. \end{aligned} \tag{4.11}$$

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

$$P_{{\rm e}|0}(h) \leq e^{-nE_{\delta}^{{\rm DS2}_1}(\lambda,\rho,J,\{\alpha_j\})}, \quad 0 \leq \rho \leq 1, \quad \lambda \geq 0, \quad \delta \triangleq \frac{h}{n}. \tag{4.12}$$

where

$$\begin{aligned} E_{\delta}^{{\rm DS2}_1}(\lambda, \rho, J, \{\alpha_j\}) \;\; &\triangleq \;\; -\rho r^{[\mathcal{C}]}(\delta) - \rho\delta \ln \left(\sum_{j=1}^{J} \alpha_j \sum_{y} \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1-\lambda\rho}{\rho}} p(y|1;j)^{\lambda} \right) \\ & -\rho(1-\delta) \ln \left(\sum_{j=1}^{J} \alpha_j \sum_{y} \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1}{\rho}} \right) \end{aligned} \tag{4.13}$$

and $r^{[\mathcal{C}]}(\delta)$ designates the normalized exponent of the distance spectrum as in (2.3).

Let

$$g_1(y;j) \triangleq p(y|0;j)^{\frac{1}{\rho}}, \quad g_2(y;j) \triangleq p(y|0;j)^{\frac{1}{\rho}} \left(\frac{p(y|1;j)}{p(y|0;j)} \right)^{\lambda} \tag{4.14}$$

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

$$\sum_y \psi(y; j) = 1, \quad j = 1 \ldots J. \tag{4.15}$$

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

$$\begin{aligned}
\psi(y; j)^{-\frac{1}{\rho}} &\left(\frac{\alpha_j (1-\delta)(1-\frac{1}{\rho}) g_1(y; j)}{\sum_y \sum_{j=1}^J \alpha_j \psi(y; j)^{1-\frac{1}{\rho}} g_1(y; j)} \right. \\
&\left. + \frac{\alpha_j \delta (1-\frac{1}{\rho}) g_2(y; j)}{\sum_y \sum_{j=1}^J \alpha_j \psi(y; j)^{1-\frac{1}{\rho}} g_2(y; j)} \right) + \xi_j = 0, \quad j = 1, \ldots, J
\end{aligned} \tag{4.16}$$

where ξ_j is a Lagrange multiplier. The solution of (4.16) is given in the following implicit form:

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

where

$$\frac{k_{2,j}}{k_{1,j}} = \frac{\delta}{1-\delta} \frac{\displaystyle\sum_{j=1}^J \sum_{y \in \mathcal{Y}} \alpha_j \psi(y; j)^{1-\frac{1}{\rho}} g_1(y; j)}{\displaystyle\sum_{j=1}^J \sum_{y \in \mathcal{Y}} \alpha_j \psi(y; j)^{1-\frac{1}{\rho}} g_2(y; j)}. \tag{4.17}$$

We note that $k \triangleq \frac{k_{2,j}}{k_{1,j}}$ in the RHS of (4.17) is *independent* of j. Thus, the substitution $\beta_j \triangleq k_{1,j}^\rho$ gives that the optimal tilting measures can be expressed as

$$\begin{aligned}
\psi(y; j) &= \beta_j \big(g_1(y; j) + k g_2(y; j)\big)^\rho \\
&= \beta_j\, p(y|0; j) \left[1 + k \left(\frac{p(y|1; j)}{p(y|0; j)}\right)^\lambda\right]^\rho \qquad y \in \mathcal{Y} \quad j = 1, \ldots, J.
\end{aligned} \tag{4.18}$$

By plugging (4.14) into (4.17) we obtain

$$k = \frac{\delta}{1-\delta} \frac{\displaystyle\sum_{j=1}^J \sum_{y \in \mathcal{Y}} \left\{ \alpha_j \beta_j^{1-\frac{1}{\rho}} p(y|0; j) \left[1 + k \left(\frac{p(y|1; j)}{p(y|0; j)}\right)^\lambda\right]^{\rho-1} \right\}}{\displaystyle\sum_{j=1}^J \sum_{y \in \mathcal{Y}} \left\{ \alpha_j \beta_j^{1-\frac{1}{\rho}} p(y|0; j) \left(\frac{p(y|1; j)}{p(y|0; j)}\right)^\lambda \left[1 + k \left(\frac{p(y|1; j)}{p(y|0; j)}\right)^\lambda\right]^{\rho-1} \right\}} \tag{4.19}$$

and from (4.14) and (4.15), β_j which is the appropriate factor normalizing the probability tilting measure $\psi(\cdot; j)$ in (4.18) is given by

$$\beta_j = \left[\sum_y p(y|0;j) \left(1 + k \left(\frac{p(y|1;j)}{p(y|0;j)} \right)^\lambda \right)^\rho \right]^{-1}, \quad j = 1, \ldots, J. \tag{4.20}$$

Note that the implicit equation for k in (4.19) and the normalization coefficients in (4.20) provide a possible generalization of the results derived in [36, 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 (4.17) 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 (4.18).

For the numerical calculation of the bound in (4.11) 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 (4.18) by first assuming an initial vector $\overline{\beta^{(0)}} = (\beta_1, \ldots, \beta_J)$ and then iterating between (4.19) and (4.20) until we get a fixed point for these equations. For a fixed δ, we need to optimize numerically the bound in (4.12) w.r.t. the two parameters λ and ρ.

4.3 Generalizing the DS2 bound to Parallel Channels: Second Approach

4.3.1 Derivation of the new bound

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. 3.2 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. 3.1, 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

$$p_B(b|x) = \alpha_j p(y|x;j) \tag{4.21}$$

because the channel state is independent of the input. If we define $-b \triangleq (-y, j)$, then we obtain from (4.21) 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. 3.2 is equivalent to an MBIOS channel with transition probability given in (4.21). 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 (4.21) into (3.15) to get the upper bound

$$\begin{aligned} P_{\mathrm{e}} \;\le\; & (A_h)^\rho \left(\sum_{j=1}^{J} \alpha_j^{\frac{1}{\rho}} \sum_y \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1-\lambda\rho}{\rho}} p(y|1;j)^{\lambda} \right)^h \\ & \left(\sum_{j=1}^{J} \alpha_j^{\frac{1}{\rho}} \sum_y \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1}{\rho}} \right)^{n-h} \qquad \begin{array}{c} 0 \le \rho \le 1 \\ \lambda \ge 0 \\ \sum_{y,j} \psi(y;j) = 1 \end{array} \end{aligned} \tag{4.22}$$

As in the first approach (see (4.9)), 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:

$$\begin{aligned} P_{\mathrm{e}} \;\le\; & (A_h)^\rho \left[\sum_{j=1}^{J} \alpha_j \sum_y g(y;j) p(y|0;j) \right]^{n(1-\rho)} \\ & \left[\sum_{j=1}^{J} \alpha_j \left(\sum_y g(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{1-\lambda} p(y|1;j)^{\lambda} \right) \right]^{h\rho} \\ & \left[\sum_{j=1}^{J} \alpha_j \left(\sum_y g(y;j)^{1-\frac{1}{\rho}} p(y|0;j) \right) \right]^{(n-h)\rho} \quad , \quad \begin{array}{c} 0 \le \rho \le 1 \\ \lambda \ge 0 \end{array} . \end{aligned} \tag{4.23}$$

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

4.3.2 Optimization of the Tilting Measures

The optimization of tilting measures for the generalized DS2 bound in (4.22) 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

$$P_{\mathrm{e}|0}(h) \le e^{-nE_\delta^{\mathrm{DS2}_2}(\lambda,\rho,J,\{\alpha_j\})}, \quad 0 \le \rho \le 1, \quad \lambda \ge 0, \quad \delta \triangleq \frac{h}{n} \tag{4.24}$$

where

$$\begin{aligned} E_\delta^{\mathrm{DS2}_2}(\lambda,\rho,J,\{\alpha_j\}) \;\triangleq\; & -\rho r^{[\mathcal{C}]}(\delta) - \rho\delta \ln\left(\sum_{j=1}^{J} \alpha_j^{\frac{1}{\rho}} \sum_y \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1-\lambda\rho}{\rho}} p(y|1;j)^{\lambda}\right) \\ & -\rho(1-\delta)\ln\left(\sum_{j=1}^{J} \alpha_j^{\frac{1}{\rho}} \sum_y \psi(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{\frac{1}{\rho}}\right). \end{aligned} \tag{4.25}$$

The optimized tilting measure should be chosen so as to maximize the exponent in (4.25). Since the perfect CSIR model is equivalent to an MBIOS channel, we can use the results of Sec. 4.2.2 with $J = 1$; by substituting the transition probability from (4.21) into (4.18), we obtain that the optimal form of the tilting measure is given by

$$\psi(y;j) = \beta\alpha_j p(y|0;j)\left(1 + k\left(\frac{p(y|1;j)}{p(y|0;j)}\right)^{\lambda}\right)^{\rho} \tag{4.26}$$

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

$$\beta = \left[\sum_{y,j} \alpha_j p(y|0;j)\left(1 + k\left(\frac{p(y|1;j)}{p(y|0;j)}\right)^{\lambda}\right)^{\rho}\right]^{-1} \tag{4.27}$$

4.4 Comparison Between the Two Generalized DS2 Bounds for Parallel Channels

Let us examine the two generalizations of the DS2 bound proposed in Sections 4.2 and 4.3 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 (4.18) with the optimal value of k in (4.19), the bound in (4.7) obtained by the first approach reads

$$\begin{aligned} P_{\mathrm{e}|0}(h) \quad &\leq \quad (A_h)^{\rho} \left\{ \left[\sum_{j=1}^{J} \alpha_j \sum_{y} \beta_j^{1-\frac{1}{\rho}} p(y|0;j) \left(\frac{p(y|1;j)}{p(y|0;j)}\right)^{\lambda} \left(1 + k_{\mathrm{opt}}^{(1)} \left(\frac{p(y|1;j)}{p(y|0;j)}\right)^{\lambda}\right)^{\rho-1} \right]^{h} \right. \\ &\left. \left[\sum_{j=1}^{J} \alpha_j \sum_{y} \beta_j^{1-\frac{1}{\rho}} p(y|0;j) \left(1 + k_{\mathrm{opt}}^{(1)} \left(\frac{p(y|1;j)}{p(y|0;j)}\right)^{\lambda}\right)^{\rho-1} \right]^{n-h} \right\}^{\rho}. \end{aligned} \tag{4.28}$$

In the same way, substituting (4.26) in (4.22) gives the bound obtained by using the second approach

$$\begin{aligned} P_{\mathrm{e}|0}(h) \quad &\leq \quad (A_h)^{\rho} \left\{ \left[\sum_{j=1}^{J} \alpha_j \sum_{y} \beta^{1-\frac{1}{\rho}} p(y|0;j) \left(\frac{p(y|1;j)}{p(y|0;j)}\right)^{\lambda} \left(1 + k_{\mathrm{opt}}^{(2)} \left(\frac{p(y|1;j)}{p(y|0;j)}\right)^{\lambda}\right)^{\rho-1} \right]^{h} \right. \\ &\left. \left[\sum_{j=1}^{J} \alpha_j \sum_{y} \beta^{1-\frac{1}{\rho}} p(y|0;j) \left(1 + k_{\mathrm{opt}}^{(2)} \left(\frac{p(y|1;j)}{p(y|0;j)}\right)^{\lambda}\right)^{\rho-1} \right]^{n-h} \right\}^{\rho}. \end{aligned} \tag{4.29}$$

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 4.3 is tighter than the one in Section 4.2. To this end, we show that the former bound attains the random coding exponent [18] 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 4.3 is derived, gives the relation

$$P_{\mathrm{e}} \leq 2^{-n(E_0(\rho)-\rho R)} \quad 0 \leq \rho \leq 1 \tag{4.30}$$

where

$$\begin{aligned} E_0(\rho) \quad &= \quad -\log_2 \left(\sum_{b} \left(\frac{1}{2} p_B(b|0)^{\frac{1}{1+\rho}} + \frac{1}{2} p_B(b|1)^{\frac{1}{1+\rho}} \right)^{1+\rho} \right) \\ &= \quad -\log_2 \left(\sum_{j=1}^{J} \alpha_j \sum_{y} \left(\frac{1}{2} p(y|0;j)^{\frac{1}{1+\rho}} + \frac{1}{2} p(y|1;j)^{\frac{1}{1+\rho}} \right)^{1+\rho} \right) \end{aligned} \tag{4.31}$$

We now turn to find the random coding exponent which stems from the use of the bound in Section 4.3. We start with the bound in (4.23) which is expressed in terms of the un-normalized tilting measure. Consider the following choice for the un-normalized tilting measure

$$g(y;j) = \left[\frac{1}{2}\,p(y|0;j)^{\frac{1}{1+\rho}} + \frac{1}{2}\,p(y|1;j)^{\frac{1}{1+\rho}}\right]^{\rho} p(y|0;j)^{-\frac{\rho}{1+\rho}} \quad , \quad j = 1, 2, \ldots, J \quad (4.32)$$

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

$$A_h = 2^{-n(1-R)}\binom{n}{h}, \qquad h = 0, 1, \ldots, n. \quad (4.33)$$

Substituting (4.32) and (4.33) into (4.9) and setting $\lambda = \frac{1}{1+\rho}$ gives the bound

$$P_{\mathrm{e}} \leq 2^{nR\rho}\left\{\sum_{j=1}^{J}\alpha_j \sum_{y}\left[\frac{1}{2}\,p(y|0;j)^{\frac{1}{1+\rho}} + \frac{1}{2}\,p(y|1;j)^{\frac{1}{1+\rho}}\right]^{1+\rho}\right\}^{n} \quad (4.34)$$

which coincides with the random coding bound in (4.30)-(4.31).

By substituting the tilting measure (4.32) in the bound in Section 4.2 (see (4.7)) we get the following error exponent, which appears instead of $E_0(\rho)$ in (4.31)

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

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 4.3 is tighter than the one in Section 4.2 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 (4.18) and (4.26) 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 (4.20)) and one constant (see (4.27)) in the second version. This suggests that neither of these bounds in (4.28) and (4.29) is uniformly tighter than the other for general codes or ensembles; this observation 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 (4.4) to (4.5)). 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.

4.5 Statement of the Main Result Derived in Chapter 4

The analysis in this chapter 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 j^{th} 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 j^{th} 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. (4.11) and (4.22) 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 (4.10) forms an upper bound on the block error probability of the whole code (or ensemble). For the bound in (4.11), 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 (4.18); for the bound in (4.22), the optimal tilting measure is given in (4.26).

Chapter 5

The Gallager Bound for Parallel Channels and Its Connection to the DS2 Bound

5.1 Short overview

The 1961 Gallager bound for a single MBIOS channel was derived in [17], and a generalization of the bound for parallel MBIOS channels was proposed by Liu et al. [26]. In the following, we outline the derivation in [26] which serves as a preliminary step towards the discussion of its relation to the two versions of the generalized DS2 bound from Chapter 4. In this chapter, 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 [26] where the authors choose some simple and sub-optimal tilting measures. By doing so, the authors in [26] 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 Chapter 8 for turbo-like ensembles).

5.2 Presentation of the Bound by Liu et al.

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

$$D_m(\underline{x}^{m'}, \underline{y}) \triangleq \ln\left(\frac{f_n^{(m)}(\underline{y})}{p_n(\underline{y}|\underline{x}^{m'})}\right) \tag{5.1}$$

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 [40], $D_m(\cdot,\cdot)$ 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

$$P_{\mathrm{e}|m} \leq \mathrm{Prob}(\underline{y} \in \mathcal{Y}_\mathrm{b}^n) + \mathrm{Prob}(\mathrm{error}, \underline{y} \in \mathcal{Y}_\mathrm{g}^n) \ . \tag{5.2}$$

We use separate bounding techniques for the two terms in (5.2). Applying the Chernoff bound on the first term gives

$$P_1 \triangleq \mathrm{Prob}(\underline{y} \in \mathcal{Y}_\mathrm{b}^n) \leq \mathbf{E}\left(e^{sW}\right), \quad s \geq 0 \tag{5.3}$$

where

$$W \triangleq \ln\left(\frac{f_n^{(m)}(\underline{y})}{p_n(\underline{y}|\underline{x}^m)}\right) - nd \ . \tag{5.4}$$

Using a combination of the union and Chernoff bounds for the second term in the RHS of (5.2) gives

$$\begin{aligned}
P_2 &\triangleq \text{Prob}(\text{error}, \underline{y} \in \mathcal{Y}_{\text{g}}^n) \\
&= \text{Prob}\left(D_m(\underline{x}^{m'}, \underline{y}) \le D_m(\underline{x}^m, \underline{y}) \text{ for some } m' \neq m,\ \underline{y} \in \mathcal{Y}_{\text{g}}^n\right) \\
&\le \sum_{m' \neq m} \text{Prob}\left(D_m(\underline{x}^{m'}, \underline{y}) \le D_m(\underline{x}^m, \underline{y}),\ D_m(\underline{x}^m, \underline{y}) \le nd\right) \\
&\le \sum_{m' \neq m} \mathbf{E}\left(\exp(tU_{m'} + rW)\right), \quad t, r \le 0 \qquad (5.5)
\end{aligned}$$

where, based on (5.1),

$$U_{m'} = D_m(\underline{x}^{m'}, \underline{y}) - D_m(\underline{x}^m, \underline{y}) = \ln\left(\frac{p_n(\underline{y}|\underline{x}^m)}{p_n(\underline{y}|\underline{x}^{m'})}\right) . \qquad (5.6)$$

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 4.2, we impose on the function $f_n^{(m)}(\underline{y})$ the restriction that it can be expressed in the product form

$$f_n^{(m)}(\underline{y}) = \prod_{j=1}^{J} \prod_{i \in \mathcal{I}(J)} f(y_i; j) . \qquad (5.7)$$

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 (4.1), (5.4), (5.6) and (5.7) into

(5.3) and (5.5) we get

$$
\begin{aligned}
P_1 &\leq \sum_{\underline{y}} \left\{ \prod_{j=1}^{J} \prod_{i \in \mathcal{I}(j)} \left(\frac{f(y_i; j)}{p(y_i|0; j)} \right)^s p(y_i|0; j) \right\} e^{-nsd} \\
&= \prod_{j=1}^{J} \left\{ \left(\sum_{y \in \mathcal{Y}} p(y|0; j)^{1-s} f(y; j)^s \right)^{n_j} \right\} e^{-nsd} \quad s \geq 0 \qquad (5.8) \\
P_2 &\leq \sum_{\underline{y}} \sum_{m' \neq m} \prod_{j=1}^{J} \prod_{i \in \mathcal{I}(j)} \left\{ \left(\frac{f(y_i; j)}{p(y_i|0; j)} \right)^r p(y_i|0; j) \left(\frac{p(y_i|0; j)}{p(y_i|x_i^{(m')}; j)} \right)^t \right\} e^{-nrd} \\
&= \sum_{h_1=0}^{n_1} \cdots \sum_{h_J=0}^{n_J} \left\{ A_{h_1, \ldots, h_J} \prod_{j=1}^{J} \left[\sum_{y \in \mathcal{Y}} p(y|0; j)^{1-r} f(y; j)^r \left(\frac{p(y|0; j)}{p(y|1; j)} \right)^t \right]^{h_j} \right. \\
&\qquad \left. \prod_{j=1}^{J} \left[\sum_{y \in \mathcal{Y}} p(y|0; j)^{1-r} f(y; j)^r \right]^{n_j - h_j} \right\} e^{-nrd}, \quad t, r \leq 0 \qquad (5.9)
\end{aligned}
$$

where as before, we use the notation $n_j \triangleq |\mathcal{I}(j)|$. Optimizing the parameter t gives the value in [17, Eq. (3.27)]

$$t = \frac{r-1}{2} \,. \qquad (5.10)$$

Let us define

$$
\begin{aligned}
G(r; j) &\triangleq \sum_{y} p(y|0; j)^{1-r} f(y; j)^r \qquad (5.11) \\
Z(r; j) &\triangleq \sum_{y} \left[p(y|0; j) p(y|1; j) \right]^{\frac{1-r}{2}} f(y; j)^r. \qquad (5.12)
\end{aligned}
$$

Substituting (5.10) into (5.9), combining the bounds on P_1 and P_2 in (5.8) and (5.9), and finally averaging over all possible channel assignments, we obtain

$$
\begin{aligned}
P_{\mathrm{e}} &\leq \mathbf{E}\left[\sum_{h=1}^{n} \sum_{\substack{0 \leq h_j \leq n_j \\ \sum h_j = h}} A_{h_1,\ldots,h_J} \prod_{j=1}^{J} [Z(r;j)]^{h_j} [G(r;j)]^{n_j - h_j} e^{-nrd} \right. \\
&\qquad\qquad \left. + \prod_{j=1}^{J} [G(s;j)]^{n_j} e^{-nsd} \right] \\
&= \sum_{\substack{n_j \geq 0 \\ \sum n_j = n}} \left\{ \sum_{h=1}^{n} \sum_{\substack{0 \leq h_j \leq n_j \\ \sum h_j = h}} A_{h_1,\ldots,h_J} \prod_{j=1}^{J} [Z(r;j)]^{h_j} [G(r;j)]^{n_j - h_j} e^{-nrd} \right. \\
&\qquad \left. + \prod_{j=1}^{J} [G(s;j)]^{n_j} e^{-nsd} \right\} P_{\underline{N}}(\underline{n}) \quad , \quad \begin{array}{c} r \leq 0 \\ s \geq 0 \\ -\infty < d < \infty \end{array} \quad . \qquad (5.13)
\end{aligned}
$$

Following the same procedure for random assignments as in (4.6) and (4.7), we obtain

$$
\begin{aligned}
P_{\mathrm{e}} &\leq \sum_{h=1}^{n} \left\{ A_h \left(\sum_{j=1}^{J} \alpha_j Z(r;j) \right)^{h} \left(\sum_{j=1}^{J} \alpha_j G(r;j) \right)^{n-h} \right\} e^{-nrd} \\
&\quad + \left(\sum_{j=1}^{J} \alpha_j G(s;j) \right)^{n} e^{-nsd} \ . \qquad (5.14)
\end{aligned}
$$

Finally, we optimize the bound in (5.14) over the parameter d which gives

$$
P_{\mathrm{e}} \leq 2^{H(\rho)} \left\{ \sum_{h=1}^{n} A_h \left[\sum_{j=1}^{J} \alpha_j Z(r;j) \right]^{h} \left[\sum_{j=1}^{J} \alpha_j G(r;j) \right]^{n-h} \right\}^{\rho} \left\{ \sum_{j=1}^{J} \alpha_j G(s;j) \right\}^{n(1-\rho)} \qquad (5.15)
$$

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

$$
\rho \triangleq \frac{s}{s-r} \quad , \quad 0 \leq \rho \leq 1 \ . \qquad (5.16)
$$

The bound in (5.15), originally derived in [26], is a natural generalization of the 1961 Gallager bound for parallel channels.

5.3 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. 4. For the case of a single MBIOS channel, it was shown [13, 37, 40] 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. 4.3 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 [13, 37, 40] therefore apply directly to the CSIR model and can be used to show that the DS2 bound in (4.22) is tighter than the 1961 Gallager bound (5.15).

In this respect, the DS2 bound from Section 4.2 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 4.2, 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 (5.8) and (5.9) 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 (5.8) and (5.9), yielding that

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

where

$$V(r,t;j) \triangleq \sum_y p(y|0;j)^{1-r} f(y;j)^r \left(\frac{p(y|0;j)}{p(y|1;j)} \right)^t \tag{5.17}$$

$G(\cdot;j)$ is introduced in (5.11) for $j = 1,\dots,J$, and ρ is given in (5.16). Averaging the bound with respect to all possible channel assignments, we get for $0 \le \rho \le 1$

$$
\begin{aligned}
P_{\mathrm{e}} &\le 2^{H(\rho)} \sum_{\substack{n_j \ge 0 \\ \sum_j n_j = n}} \Bigg\{ \Bigg[\sum_{h=1}^{n} \sum_{\substack{h_1,\dots,h_j \\ \sum_j h_j = h}} A_{h_1,\dots,h_j} \prod_{j=1}^{J} \left[V(r,t;j) G(s;j)^{\frac{1-\rho}{\rho}} \right]^{h_j} \\
&\qquad \prod_{j=1}^{J} \left[G(r;j) G(s;j)^{\frac{1-\rho}{\rho}} \right]^{n_j - h_j} \Bigg]^{\rho} P_{\underline{N}}(\underline{n}) \Bigg\} \\
&\le 2^{H(\rho)} \Bigg[\sum_{\substack{n_j \ge 0 \\ \sum_j n_j = n}} \sum_{h=1}^{n} \sum_{\substack{h_1,\dots,h_j \\ \sum_j h_j = h}} A_{h_1,\dots,h_j} P_{\underline{N}}(\underline{n}) \prod_{j=1}^{J} \left[V(r,t;j) G(s;j)^{\frac{1-\rho}{\rho}} \right]^{h_j} \\
&\qquad \prod_{j=1}^{J} \left[G(r;j) G(s;j)^{\frac{1-\rho}{\rho}} \right]^{n_j - h_j} \Bigg]^{\rho} \qquad (5.18)
\end{aligned}
$$

where we invoked Jensen's inequality in the last step. Following the same steps as in (4.4)–(4.7), we get

$$
\begin{aligned}
P_{\mathrm{e}} &\le 2^{H(\rho)} \Bigg[\sum_{h=1}^{n} A_h \left(\sum_{j=1}^{J} \alpha_j V(r,t;j) G(s;j)^{\frac{1-\rho}{\rho}} \right)^{h} \\
&\qquad \left(\sum_{j=1}^{J} \alpha_j G(r;j) G(s;j)^{\frac{1-\rho}{\rho}} \right)^{n-h} \Bigg]^{\rho}, \qquad (5.19)
\end{aligned}
$$

where from (5.10), (5.11), (5.16) and (5.17)

$$
\begin{aligned}
G(s;j) &= \sum_y p(y|0;j) \left(\frac{f(y;j)}{p(y|0;j)} \right)^{s} \\
G(r;j) &= \sum_y p(y|0;j) \left(\frac{f(y;j)}{p(y|0;j)} \right)^{s(1-\frac{1}{\rho})} \\
V(r,t;j) &= \sum_y p(y|0;j) \left(\frac{f(y;j)}{p(y|0;j)} \right)^{s(1-\frac{1}{\rho})} \left(\frac{p(y|0;j)}{p(y|1;j)} \right)^{t}. \qquad (5.20)
\end{aligned}
$$

Setting $\lambda = -t$, and substituting in (5.20) the following relation between the Gallager tilting measures and the un-normalized tilting measures in the DS2 bound

$$
g(y;j) \triangleq \left(\frac{f(y;j)}{p(y|0;j)} \right)^{s}, \quad j = 1,2,\dots,J \qquad (5.21)
$$

we obtain

$$
\begin{aligned}
P_{\mathrm{e}} \;\le\; & 2^{H(\rho)} \left\{ \sum_{h=0}^{n} A_h \left[\sum_{j=1}^{J} \alpha_j \left(\sum_y g(y;j)^{1-\frac{1}{\rho}} p(y|0;j)^{1-\lambda} p(y|1;j)^{\lambda} \right) \right.\right. \\
& \left. \left(\sum_y g(y;j) p(y|0;j) \right)^{\frac{1-\rho}{\rho}} \right]^{h} \left[\sum_{j=1}^{J} \alpha_j \left(\sum_y g(y;j)^{1-\frac{1}{\rho}} p(y|0;j) \right) \right. \\
& \left.\left. \left(\sum_y g(y;j) p(y|0;j) \right)^{\frac{1-\rho}{\rho}} \right]^{n-h} \right\}^{\rho} , \quad 0 \le \rho \le 1 \qquad (5.22)
\end{aligned}
$$

which coincides with the form of the DS2 bound given in (4.9) (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 (5.21) are *even*.

Discussion. The derivation of the 1961 Gallager bound first involves the averaging of the bound in (5.13) over all possible channel assignments and then the optimization over the parameter d in (5.14). To show a connection to the DS2 bound in (4.9), 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 (5.18) 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 [26], the need for Jensen's inequality was circumvented due to the linearity of the expression in (5.13). We note that Jensen's inequality was also used for the direct derivation of the DS2 bound in (4.7); 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. 3.4. In this case, as in the case of the DS2 bound from Section 4.3, the DS2 bound is tighter than the 1961 Gallager bound (as noted in [40]) 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 4 because the optimal choice of $\psi(\cdot;j)$ given in (4.18) 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.

- 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 (4.9) 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).

5.4 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 one of them also includes the all-zero codeword). The 1961 Gallager bound is applied separately for every subcode, and the union bound (4.10) is taken over the subcodes. Using these optimized tilting measures improves the tightness of the resulting bound, as exemplified in the continuation of this dissertation.

5.4.1 Tilting Measures for Random Codes

Consider the ensemble of fully random binary block codes of length n. Substituting the appropriate weight enumerator (given in (2.5)) into (5.14), we get

$$\begin{aligned} P_{\mathrm{e}} \;\leq\; & 2^{-n(1-R)} \left\{ \frac{1}{2} \sum_{j=1}^{J} \alpha_j \sum_{y} \left[p(y|0;j)^{\frac{1-r}{2}} + p(y|1;j)^{\frac{1-r}{2}} \right]^2 f(y;j)^r \right\}^n e^{-nrd} \\ & + \left\{ \frac{1}{2} \sum_{j=1}^{J} \alpha_j \sum_{y} \left(p(y|0;j)^{1-s} + p(y|1;j)^{1-s} \right) f(y;j)^s \right\}^n e^{-nsd}\,, \quad \begin{array}{l} r \leq 0 \\ s \geq 0 \\ d \in \mathbb{R} \end{array} \end{aligned} \tag{5.23}$$

where we rely on (5.11) and (5.12), 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 (5.23) 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 $L \in \mathbb{R}$. This equation is satisfied for functions which are given in the form

$$f(y;j) = K \left\{ \frac{\left(p(y|0;j)^{\frac{1-r}{2}} + p(y|1;j)^{\frac{1-r}{2}} \right)^2}{p(y|0;j)^{1-s} + p(y|1;j)^{1-s}} \right\}^{\frac{1}{s-r}} \qquad K \in \mathbb{R}. \tag{5.24}$$

This forms a natural generalization of the tilting measure given in [17, 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 (5.24) in (5.15).

5.4.2 Tilting Measures for Constant Hamming Weight Codes

The distance spectrum of a constant Hamming weight code is given by

$$A_{h'} = \begin{cases} 1, & \text{if } h' = 0 \\ A_h, & \text{if } h' = h \\ 0, & \text{otherwise} \end{cases} \tag{5.25}$$

Substituting this into (5.15) 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

$$\begin{aligned}
P_{\mathrm{e}|0}(h) \quad &\leq \quad 2^{H(\rho)} A_h^{\rho} \left\{ \sum_{j=1}^{J} \alpha_j \sum_y \left[p(y|0;j) p(y|1;j) \right]^{\frac{1-r}{2}} f(y;j)^r \right\}^{h\rho} \\
&\cdot \left\{ \sum_{j=1}^{J} \frac{\alpha_j}{2} \sum_y \left[p(y|0;j)^{1-r} + p(y|1;j)^{1-r} \right] f(y;j)^r \right\}^{(n-h)\rho} \\
&\cdot \left\{ \sum_{j=1}^{J} \frac{\alpha_j}{2} \sum_y \left[p(y|0;j)^{1-s} + p(y|1;j)^{1-s} \right] f(y;j)^s \right\}^{n(1-\rho)} , \\
&\qquad r \leq 0,\ s \geq 0,\ \rho = \frac{s}{s-r}. \qquad (5.26)
\end{aligned}$$

Applying calculus of variations to (5.26) yields (see Appendix B for some additional details) that the following condition should be satisfied for all values of $y \in \mathcal{Y}$:

$$\begin{aligned}
\sum_{j=1}^{J} \alpha_j \Big\{ \left[p(y|0;j)^{1-s} + p(y|1;j)^{1-s} \right] f(y;j)^{s-r} + K_1 \left[p(y|0;j) p(y|1;j) \right]^{\frac{1-r}{2}} \qquad (5.27) \\
+ K_2 \left[p(y|0;j)^{1-r} + p(y|1;j)^{1-r} \right] \Big\} \quad = \quad 0
\end{aligned}$$

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

$$\begin{aligned}
f(y;j) \quad = \quad &\left\{ \frac{c_1 \left(p(y|0;j)^{\frac{1-s(1-\rho^{-1})}{2}} + p(y|1;j)^{\frac{1-s(1-\rho^{-1})}{2}} \right)^2}{p(y|0;j)^{1-s} + p(y|1;j)^{1-s}} + \right. \\
&\left. \frac{d_1 \ \left(p(y|0;j)^{1-s(1-\rho^{-1})} + p(y|1;j)^{1-s(1-\rho^{-1})} \right)}{p(y|0;j)^{1-s} + p(y|1;j)^{1-s}} \right\}^{\frac{\rho}{s}} , \quad \begin{array}{c} c_1, d_1 \in \mathbb{R} \\ s \geq 0 \\ 0 \leq \rho \leq 1 \end{array} \qquad (5.28)
\end{aligned}$$

where we have used (5.16). 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

(5.28) in (5.15)). 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 (5.28) to finite ones. It is shown in [35, 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 [35, Appendix B], enables one to express the optimized tilting measure in (5.28) using an equivalent form where the new parameter c ranges in the interval $[0,1]$. The numerical optimization of the bound in (5.28) 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 [35, Appendices A, B], the functions $f(\cdot; j)$ get the equivalent form

$$\begin{aligned} f(y;j) \quad = \quad & \left\{ \frac{(1-c)\left(p(y|0;j)^{\frac{1-s(1-\rho^{-1})}{2}} - p(y|1;j)^{\frac{1-s(1-\rho^{-1})}{2}}\right)^2}{p(y|0;j)^{1-s} + p(y|1;j)^{1-s}} \right. \\ & \left. + \frac{2c\left(p(y|0;j)p(y|1;j)\right)^{\frac{1-s(1-\rho^{-1})}{2}}}{p(y|0;j)^{1-s} + p(y|1;j)^{1-s}} \right\}^{\frac{\rho}{s}}, \quad (\rho, s, c) \in [0,1]^3. \quad (5.29) \end{aligned}$$

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

5.5 Alternative Derivation of the 1961 Gallager Bound Using the CSIR Approach

In this section we use the CSIR approach to obtain a generalization of the 1961 Gallager bound for parallel channels, where the application of this approach is very similar to that of Sec. 4.3. The key result here is that applying the CSIR approach to the 1961 Gallager bound yields the same bound as in (5.15). The derivation is as follows. We begin with the classic 1961 Gallager bound [17] for a single MBIOS channel given that the all-zero codeword is transmitted. Applying the bound on a constant-Hamming weight subcode yields

$$P_{\mathrm{e}|0}(h) \leq 2^{H(\rho)} \left\{ A_h \left[\sum_y [p(y|0)p(y|1)]^{\frac{1-r}{2}} \tilde{f}(y)^r \right] \left[\sum_y p(y|0)^{1-r} \tilde{f}(y)^r \right]^{n-h} \right\}^{\rho} \left\{ \sum_y p(y|0)^{1-s} \tilde{f}(y)^s \right\}^{n(1-\rho)} \tag{5.30}$$

where $\tilde{f}(\cdot)$ is the tilting measure associated with the bound. The optimal tilting measure for this bound is of the form

$$\tilde{f}(y) = \left\{ \frac{(1-c)\left(p(y|0)^{\frac{1-r}{2}} + p(y|1)^{\frac{1-r}{2}}\right)^2 + 2c\left(p(y|0)^{1-r} + p(y|1)^{1-r}\right)}{p(y|0)^{1-s} + p(y|1)^{1-s}} \right\}^{\frac{\rho}{s}} \tag{5.31}$$

for some $\rho, s, c \in [0,1]^3$. We now make use of the CSIR model by replacing y with $b = (y, j)$ and substituting (4.21) in (5.30) and (5.31); this yields the bound

$$P_{\mathrm{e}|0}(h) \leq 2^{H(\rho)} \left\{ A_h \left[\sum_{j=1}^{J} \sum_y \alpha_j^{1-r} [p(y|0;j)p(y|1;j)]^{\frac{1-r}{2}} \tilde{f}(y;j)^r \right]^h \left[\sum_{j=1}^{J} \sum_y \alpha_j^{1-r} p(y|0;j)^{1-r} \tilde{f}(y;j)^r \right]^{n-h} \right\}^{\rho} \left\{ \sum_y p(y|0;j)^{1-s} \tilde{f}(y;j)^s \right\}^{n(1-\rho)} \tag{5.32}$$

and the optimal tilting measure

$$\tilde{f}(y;j) = \alpha_j f(y;j) \tag{5.33}$$

where $f(y;j)$ is given in (5.28). By substituting (5.33) in (5.32) the optimal bound is

$$P_{\mathrm{e}|0}(h) \leq 2^{H(\rho)} \left\{ A_h \left[\sum_{j=1}^{J} \alpha_j \sum_y [p(y|0;j)p(y|1;j)]^{\frac{1-r}{2}} f(y;j)^r \right] \left[\sum_{j=1}^{J} \alpha_j \sum_y p(y|0;j)^{1-r} f(y;j)^r \right]^{n-h} \right\}^{\rho} \left\{ \sum_{j=1}^{J} \alpha_j \sum_y p(y|0;j)^{1-s} f(y;j)^s \right\}^{n(1-\rho)} \tag{5.34}$$

which is identical to the 1961 Gallager bound in (5.15), when applied for a constant Hamming-weight code.

5.6 Statement of the Main Result Derived in Chapter 5

The analysis in this chapter 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 (5.15) provides an upper bound on the ML decoding error probability when the bound is taken over the whole code (as originally derived in [26]). 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 (5.26), and (4.10) 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 (5.26), is given by (5.29); this set of functions is subject to a three-parameter optimization over a cube of unit length (see (5.29)).

Chapter 6

Special Cases of the Generalized DS2 Bound for Independent Parallel Channels

6.1 Short overview

In this chapter, we rely on the two versions of the generalized DS2 bound for independent parallel MBIOS channels, as presented in Sections 4.2 and 4.3, and apply them in order to re-derive some of the bounds which were originally derived by Liu et al. [26]. The derivation in [26] is based on the 1961 Gallager bound from Section 5.2, 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 5.4 which lead to more complicated bounds in terms of their numerical computation). In this chapter, 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 [26] as special cases of the generalized DS2 bound, or alternatively, obtain some modified bounds as compared to [26].

6.2 The Union-Bhattacharyya Bound

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

$$P_{\mathrm{e}} \leq \sum_{h=1}^{n(1-R)} A_h \gamma^h \tag{6.1}$$

where γ is given by (3.3) 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, we now present a tighter version which uses the Q-function to express the exact pairwise error probability.

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. 4. This is because when applying the perfect 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

$$P_{\mathrm{e}}(\underline{0} \to \underline{x}_{h_1,h_2,\ldots,h_J}) = Q\left(\sqrt{2\sum_{j=1}^{J} \nu_j h_j}\right) \tag{6.2}$$

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 $\nu_j \triangleq \left(\frac{E_{\mathrm{s}}}{N_0}\right)_j$ designates the energy per symbol to spectral noise density for the j^{th} AWGN channel ($j = 1, 2, \ldots, J$). The union bound on the block error probability gives

$$P_{\mathrm{e}} \;\leq\; \sum_{h=1}^{n} \sum_{\substack{h_1 \geq 0, \ldots, h_J \geq 0 \\ h_1 + \ldots + h_J = h}} A_{h_1,\ldots,h_J} Q\left(\sqrt{2\sum_{j=1}^{J} \nu_j h_j}\right) \tag{6.3}$$

where this bound is expressed in terms of the split weight enumerator of the code. Averaging (6.3) over all possible channel assignments gives (see (4.6))

$$
\begin{aligned}
P_{\mathrm{e}} \;&\leq \sum_{\substack{n_j \geq 0 \\ n_1+\ldots+n_J=n}} \left\{ \sum_{h=1}^{n} \sum_{\substack{0 \leq h_j \leq n_j \\ \sum_j h_j = h}} A_h \; P_{\underline{H}|\underline{N}}(\underline{h}|\underline{n}) \; P_{\underline{N}}(\underline{n}) \; Q\left(\sqrt{2 \sum_{j=1}^{J} \nu_j h_j}\right) \right\} \\
&= \sum_{\substack{n_j \geq 0 \\ n_1+\ldots+n_J=n}} \left\{ \sum_{h=1}^{n} \sum_{\substack{0 \leq h_j \leq n_j \\ \sum_j h_j = h}} A_h \binom{h}{h_1, h_2, \ldots, h_J} \right. \\
&\qquad\qquad \binom{n-h}{n_1-h_1, n_2-h_2, \ldots, n_J-h_J} \\
&\qquad\qquad \left. \alpha_1^{n_1} \ldots \alpha_J^{n_J} Q\left(\sqrt{2 \sum_{j=1}^{J} \nu_j h_j}\right) \right\}
\end{aligned}
\tag{6.4}
$$

where α_j designates the a-priori probability for the transmission of symbols over the j^{th} 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, i.e.,

$$
Q(x) = \frac{1}{\pi} \int_0^{\frac{\pi}{2}} e^{-\frac{x^2}{2\sin^2\theta}} \mathrm{d}\theta \,, \qquad x \geq 0. \tag{6.5}
$$

Plugging (6.5) into (6.4) and interchanging the order of integration and summation gives

$$
\begin{aligned}
P_{\mathrm{e}} \;&\le\; \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \sum_{\substack{n_j\ge 0\\ n_1+\ldots+n_J=n}} \left\{ \sum_{h=1}^{n} \sum_{\substack{0\le h_j\le n_j\\ \sum h_j=h}} A_h \binom{h}{h_1,h_2,\ldots,h_J} \right.\\
&\qquad \left. \binom{n-h}{n_1-h_1,n_2-h_2,\ldots,n_J-h_J} \alpha_1^{n_1}\ldots\alpha_J^{n_J} \prod_{j=1}^{J} e^{-\frac{\nu_j h_j}{\sin^2\theta}} \right\} \mathrm{d}\theta\\
&\overset{\text{(a)}}{=}\; \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \sum_{h=1}^{n} A_h \sum_{\substack{h_j\ge 0\\ \sum h_j=h}} \left\{ \binom{h}{h_1,h_2,\ldots,h_J} \prod_{j=1}^{J} \left[\alpha_j e^{-\frac{\nu_j}{\sin^2\theta}}\right]^{h_j} \right\}\\
&\qquad \sum_{\substack{k_j\ge 0\\ \sum_j k_j=n-h}} \left\{ \binom{n-h}{k_1,k_2,\ldots,k_J} \prod_{j=1}^{J} (\alpha_j)^{k_j} \right\} \mathrm{d}\theta\\
&\overset{\text{(b)}}{=}\; \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \sum_{h=1}^{n} A_h \left[\sum_{j=1}^{J} \alpha_j e^{-\frac{\nu_j}{\sin^2\theta}}\right]^{h} \mathrm{d}\theta
\end{aligned}
\tag{6.6}
$$

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

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

Eq. (6.6) provides the exact (Q-form) version of the union bound on the block error probability for independent parallel AWGN channels.

6.3 The Simplified 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 [26], 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 (5.21), 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

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

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

$$\begin{aligned}
\int_{-\infty}^{+\infty} g(y;j)p(y|0;j)dy &= \sqrt{\frac{1-\rho}{1-\beta\rho}} \\
\int_{-\infty}^{+\infty} g(y;j)^{1-\frac{1}{\rho}}p(y|0;j)dy &= \beta^{-\frac{1}{2}} \\
\int_{-\infty}^{+\infty} g(y;j)^{1-\frac{1}{\rho}}p(y|0;j)^{1-\lambda}p(y|1;j)^{\lambda}dy &= \frac{\gamma_j^{\beta}}{\sqrt{\beta}}\,, \quad \gamma_j \triangleq e^{-\nu_j}.
\end{aligned} \tag{6.7}$$

Next, by plugging (6.7) into the DS2 bound in (4.9), we get

$$P_{\mathrm{e}} \le \left\{\sum_{h=0}^{n} A_h \left(\sum_{j=1}^{J}\alpha_j\gamma_j^{\beta}\right)^h \beta^{-\frac{n}{2}}\right\}^{\rho} \left(\frac{1-\rho}{1-\beta\rho}\right)^{\frac{n(1-\rho)}{2}}\,, \quad \begin{array}{c} 0 \le \rho \le 1 \\ 1 \le \beta \le \frac{1}{\rho} \end{array}. \tag{6.8}$$

The same expression may be obtained by plugging (6.7) into the DS2 bound in (4.23). This bound is identical to the parallel-channel simplified sphere bound in [26, Eq. (24)], except that it provides a slight improvement due to the absence of the factor $2^{H(\rho)}$ which appears in [26, Eq. (24)] (a factor bounded between 1 and 2).

6.4 Generalizations of the Shulman-Feder Bound for Parallel Channels

In this 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. [26] 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 4.3, and the second bound follows as a particular case of the DS2 bound from Section 4.2 for independent parallel channels.

By substituting in (5.15) the tilting measure and the parameters (see [26, Eq. (28)])

$$\begin{aligned} f(y;j) &= \left(\frac{1}{2}p(y|0;j)^{\frac{1}{1+\rho}} + \frac{1}{2}p(y|1;j)^{\frac{1}{1+\rho}}\right)^{1+\rho} \\ r &= -\frac{1-\rho}{1+\rho}, \quad s = \frac{\rho}{1+\rho}, \qquad 0 \le \rho \le 1 \end{aligned} \tag{6.9}$$

straightforward calculations for MBIOS channels give the following bound which was originally introduced in [26, Lemma 2]:

$$P_{\mathrm{e}} \le 2^{H(\rho)} 2^{nR\rho} \left(\max_{1 \le h \le n} \frac{A_h}{2^{-n(1-R)}\binom{n}{h}}\right)^{\rho} \left\{ \sum_{j=1}^{J} \alpha_j \left(\sum_y \frac{1}{2}p(y|0;j)^{\frac{1}{1+\rho}} + \frac{1}{2}p(y|1;j)^{\frac{1}{1+\rho}}\right)^{1+\rho} \right\}^{n}. \tag{6.10}$$

Due to the natural connection between the DS2 bound in Section 4.3 and the 1961 Gallager bound for parallel channels (see the discussion in Sec. 5.3), 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 (4.32) (as a part of the discussion Sec. 4.4 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 (6.10) except for the absence of the factor $2^{H(\rho)}$.

Considering the generalization of the DS2 bound in Section 4.2, it is possible to start from Eq. (4.9) and take the maximum distance spectrum term out of the sum.

This gives the bound

$$
\begin{aligned}
P_{\mathrm{e}} &\le 2^{-n(1-R)\rho} \left(\max_{1 \le h \le n} \frac{A_h}{2^{-n(1-R)} \binom{n}{h}} \right)^{\rho} \left\{ \sum_{j=1}^{J} \alpha_j \left[\sum_y g(y;j) p(y|0;j) \right]^{\frac{1-\rho}{\rho}} \right. \\
& \left. \cdot \left[\sum_y p(y|0;j) g(y;j)^{1-\frac{1}{\rho}} \left(1 + \left(\frac{p(y|1;j)}{p(y|0;j)} \right)^{\lambda} \right) \right] \right\}^{n\rho}, \quad 0 \le \rho \le 1. \qquad (6.11)
\end{aligned}
$$

Using the J un-normalized tilting measures from (4.32) and setting $\lambda = \frac{1}{1+\rho}$ in (6.11), gives the following bound due to the symmetry at the channel outputs:

$$
\begin{aligned}
P_{\mathrm{e}} &\le 2^{nR\rho} \left(\max_{1 \le h \le n} \frac{A_h}{2^{-n(1-R)} \binom{n}{h}} \right)^{\rho} \\
& \left\{ \sum_{j=1}^{J} \alpha_j \left[\left(\sum_y \frac{1}{2} p(y|0;j)^{\frac{1}{1+\rho}} + \frac{1}{2} p(y|1;j)^{\frac{1}{1+\rho}} \right)^{1+\rho} \right]^{\frac{1}{\rho}} \right\}^{n\rho}, \quad 0 \le \rho \le 1 \quad (6.12)
\end{aligned}
$$

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 (6.10). The fact that the bound in (6.12) is exponentially looser than the bound in (6.10) follows from the use of Jensen's inequality for the derivation of the first version of the DS2 bound (see the move from (4.4) to (4.5)).

6.5 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

$$
\max_{1 \le h \le n} \frac{A_h}{2^{-n(1-R)} \binom{n}{h}} \qquad (6.13)
$$

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. 2.5 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 [29]. Typically, the set Ψ_n^+ consists of low and high Hamming weights, where the ratio in (6.13) 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 [26, Lemma 3], and gets the form

$$\begin{aligned} P_{\mathrm{e}} \ \leq \ & \sum_{h:\ \frac{h}{n}\in\Psi_n^+} A_h \gamma^h + 2^{H(\rho)} 2^{nR\rho} \left(\max_{h:\ \frac{h}{n}\in\Psi_n^-} \frac{A_h}{2^{-n(1-R)}\binom{n}{h}} \right)^{\rho} \\ & \cdot \left\{ \sum_{j=1}^{J} \alpha_j \left(\sum_y \frac{1}{2} p(y|0;j)^{\frac{1}{1+\rho}} + \frac{1}{2} p(y|1;j)^{\frac{1}{1+\rho}} \right)^{1+\rho} \right\}^n \end{aligned} \tag{6.14}$$

where γ is introduced in (3.3), 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 [26, Eq. (42)])

$$\delta \in \begin{cases} \Psi_n^+ & \text{if} \quad -\delta \ln \gamma \geq H(\delta) + (\overline{I} - 1) \ln 2 \\ \Psi_n^- & \text{otherwise} \end{cases} \tag{6.15}$$

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 (6.14) 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 (6.6), thus providing the following tightened upper bound:

$$
\begin{aligned}
P_{\mathrm{e}} \ \leq \ & \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \left\{ \sum_{h:\ \frac{h}{n} \in \Psi_n^+} A_h \left[\sum_{j=1}^{J} \alpha_j e^{-\frac{\nu_j}{\sin^2\theta}} \right]^h \mathrm{d}\theta \right\} \\
& + 2^{H(\rho)} 2^{nR\rho} \left(\max_{h:\ \frac{h}{n} \in \Psi_n^-} \frac{A_h}{2^{-n(1-R)} \binom{n}{h}} \right)^{\rho} \\
& \cdot \left\{ \sum_{j=1}^{J} \alpha_j \left(\sum_y \frac{1}{2} p(y|0;j)^{\frac{1}{1+\rho}} + \frac{1}{2} p(y|1;j)^{\frac{1}{1+\rho}} \right)^{1+\rho} \right\}^n . \qquad (6.16)
\end{aligned}
$$

On the selection of a suitable partitioning of the set Ψ_n in (6.16): The asymptotic partitioning suggested in (6.15) 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 (6.15) 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 (6.14)). 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 (6.14) and (6.16). 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 RHS of (6.14) and (6.16) 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 (6.15). 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 (6.16).

This algorithm is applied to the calculation of the LMSF bound for finite-length codes (see, e.g., Fig. 8.2 in p. 93).

Clearly, an alternative and slightly tighter version of the MSF bound can be obtained from the DS2 bound from Section 4.3 for parallel channels where the difference will be in the absence of the $2^{H(\rho)}$ constant. We address the MSF bound in Chapter 8, where for various ensembles of turbo-like codes, its tightness is compared with that of both versions of generalized DS2 and Gallager bounds.

Chapter 7

Inner Bounds on Attainable Channel Regions

7.1 Short overview

In this chapter, 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 4 and 5 (see Theorems 1 and 2), referring here to the asymptotic case where we let the block length tend to infinity.

7.2 Bounds on Attainable Channel Regions

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 [26], we evaluate inner bounds on the attainable channel regions whose calculation is based on upper bounds on the ML decoding error probability.

In [26, 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 [26], 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. $$\inf_{\delta_0 < \delta \leq 1} E^{\mathrm{DS2}_1}(\delta) > 0, \quad \forall\ \delta_0 \in (0,1) \tag{7.1}$$

   where, for $0 < \delta \leq 1$, $E^{\mathrm{DS2}_1}(\delta)$ is calculated from (4.13) 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

   $$\limsup_{\delta \to 0} \frac{r^{[\mathcal{C}]}(\delta)}{\delta} < -\ln\left(\sum_{j=1}^{J} \alpha_j \gamma_j\right) \tag{7.2}$$

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 (2.3).

3. There exists a sequence $\{D_n\}$ of natural numbers tending to infinity with increasing n so that
$$\limsup_{n\to\infty} \sum_{h=1}^{D_n} A_h^{[\mathcal{C}(n)]} = 0 \tag{7.3}$$

4. The normalized exponent of the distance spectrum satisfies
$$r^{[\mathcal{C}(n)]}(\delta) = r^{[\mathcal{C}]}(\delta) + o\left(\frac{D_n}{n}\right),$$
i.e., $r^{[\mathcal{C}(n)]}(\delta)$ converges uniformly in $\delta \in [0,1]$ to $r^{[\mathcal{C}]}(\delta)$ at a fast enough rate.

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 C.

Discussion: We note that conditions 3 and 4 in Theorem 3 are similar to the last two conditions in [25, Theorem 2.3]. Condition 2 above happens to be a natural generalization of the second condition in [25, Theorem 2.3], thus generalizing the single channel case to a set of parallel channels. The distinction between [25, 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 involve the generalized 1961 Gallager bound for parallel channels and the generalized DS2 bound from Section 4.3 can be proven in the same way by replacing the first condition with an equivalent relation involving the exponent of these bounds maximized over their respective parameters, instead of the error exponent of the DS2 bound from Section 4.2.

The difference of our results from those presented in [26] stems from the fact that we rely here on the generalized DS2 bounds 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 [26, 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 [11, 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 $D_n = O\left(\ln(n)\right)$ (so the sequence $\{D_n\}$ tends to infinity logarithmically with n). Based on the calculations of the distance spectrum of this ensemble (see [11, 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

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

Hence, inequality (7.2) in Theorem 3 (i.e., the second condition in this theorem) is also satisfied for this ensemble (since the RHS of (7.2) 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 (7.1)).

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 [17, Section 2.2]), so the LHS of (7.2) becomes negative.

We make use of the fulfillment of the condition in (7.2) for regular NSRA codes and some other variants of accumulate-based codes later in Section 8.3.

It is important to note that the low Hamming weight codewords which are addressed by the requirement in (7.3) may yield that the error probability under ML decoding does not necessarily vanish exponentially with the block length (see, e.g., [29, Theorems 3 and 4] and [11, 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).

Chapter 8

Performance Bounds for Turbo-Like Ensembles over Parallel Channels

8.1 Overview

In this chapter, 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 [26], showing the superiority of the new bounds introduced in Chapters 4 and 5. As mentioned before, the superiority of the generalized 1961 Gallager bound in Chapter 5 over the LMSF bound from [26] is attributed to the optimization of its related tilting measures.

We focus especially on ensembles of accumulate-based codes presented in Chapter 2, i.e, uniformly interleaved ensembles of repeat-accumulate (RA) and accumulate-repeat-accumulate (ARA) codes. These codes, originally introduced by Divsalar et al. [2, 11], are attractive since they possess low encoding and decoding complexity under iterative decoding and show a remarkable improvement in performance over classical algebraic codes. For independent parallel channels, we study their theoretical performance under ML decoding and compare it to their performance under iterative decoding. Both finite-length analysis and asymptotic analysis are considered. 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.

The structure of this chapter is as follows. Section 8.2 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 8.3 focuses on performance bounds for repeat-accumulate codes and their recent variations which were presented in Chapter 2. The attractiveness of these ensembles is due to their remarkable performance and low encoding and decoding complexity under iterative decoding algorithms. 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. These results are compared with computer simulations of suboptimal iterative decoding. In Section 8.4 we discuss 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.

8.2 Performance Bounds for Uniformly Interleaved Turbo Codes

In this 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 for turbo code ensembles refer to parallel BIAWGN channels. The reader is referred to [24] 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. 8.2 compares upper bounds on the bit error probability of the ensemble of

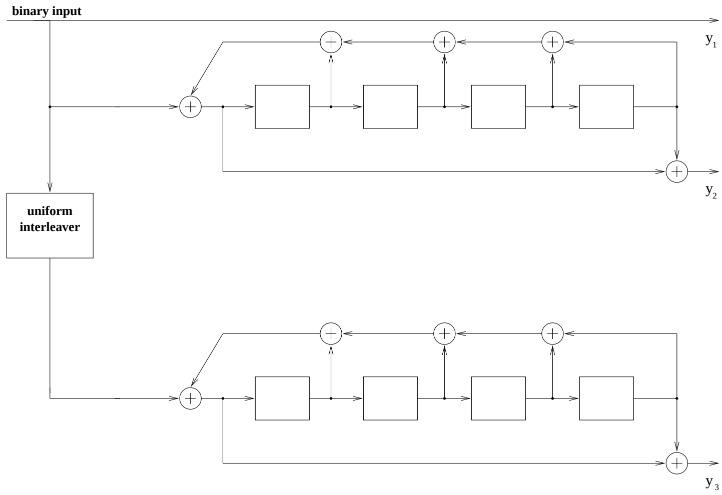


Figure 8.1: The encoder of an ensemble of uniformly interleaved turbo codes whose interleaver is of length 1000, and there is no puncturing of parity bits.

uniformly interleaved turbo codes of rate $R = \frac{1}{3}$ bits per channel use (see Fig. 8.2). 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 [28] for the calculation of the IOWE of convolutional codes), and finally, the uniform interleaver which is placed between the two constituent codes in Fig. 8.2 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 $\left(\frac{E_\mathrm{b}}{N_0}\right)_1 = 0$ 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 $\left(\frac{E_\mathrm{b}}{N_0}\right)_2$ below the value in the RHS of (3.8) (whose value in this setting is 3.69 dB). This limitation of the union bound is indeed reflected from Fig. 8.2, thus showing how loose is the union bound as compared to computer simulations of the (sub-optimal) iterative decoder. The LMSF bound depicted in Fig. 8.2 uses a partitioning for codes of finite length which was obtained via the algorithm described in Section 6.5; 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.

8.3 Performance Bounds for Ensembles of Accumulate-Based Codes

In 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. 2.3: 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 (SPRA) codes and ARA (SPARA) codes, respectively, where the number of repetitions is equal to $q = 6$ and, as a result of puncturing, only 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) which were introduced in Section 2.5. The calculation of the IOWEs of these three ensembles is performed in Section 2.5 and we rely on the results of this analysis in the evaluation of inner bounds on attainable channel regions. The two generalizations of the DS2 bound for parallel channels are then applied to these ensembles for the asymptotic case where we let the block length tend to infinity.

The evaluation of inner bounds on the attainable channel regions for the considered ensembles of accumulate-based codes in this section is based on Theorem 3.

In Fig. 8.3, we compare inner bounds on the attainable channel boundaries as calculated by the union, LMSF, and DS2 bounds from Sections 4.2 and 4.3. This plot refers to the ensemble of NSRA($N, 3$) codes of rate $\frac{1}{3}$ bits per channel use (see Fig. 2.3 (a)) where we let N tend to infinity. The asymptotic growth rate of the distance spectrum of this ensemble is calculated by (2.23) 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 Section 4.3 appears to be slightly tighter than the DS2 bound from Section 4.2 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 (4.4) to (4.5)) hinders its tightness the most, possibly due to the large variance of the summands in (4.4). This phenomenon was also observed for various turbo-like ensembles, 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 Section 4.2 is in some cases tighter than the DS2 bound from Section 4.3. 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. 8.4.

This figure demonstrates the improved performance of the ensembles of SPARA codes under ML decoding. This improvement is attributed to the distance spectral thinning effect [30] which is exemplified in Fig. 2.5 for the ensembles of NSRA, SPRA and SPARA codes of the same code rate ($\frac{1}{3}$ bits per channel use) where we can see the resemblance between the distance spectrum of these ensembles to that of the random code ensemble. The same phenomenon of distance spectral thinning occurs by reducing the value of α for the ensembles of SPARA codes (see Fig. 2.5, 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. 8.4. 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. 8.4).

While these results hold for the case of ML decoding, it is clearly of interest to examine the performance of these code ensembles under iterative decoding. A comparison of the performance of these ensembles is given in Fig. 8.5. In this figure, the performance of some SPARA codes of rate $\frac{1}{3}$ is obtained via computer simulations employing 32 iterations of the sum-product decoding algorithm. In this figure, the transmission takes place over two parallel binary-input AWGN channels, where 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, and each bit is equally likely to be assigned to one of these two channels. An external high-rate code is used which improves performance at the cost of slightly reducing the coding rate. It is apparent from Fig. 8.5 that in the case of iterative decoding, the optimal value of α lies between $\frac{1}{4}$ and $\frac{1}{3}$. This is in contrast to the performance bounds for ML decoding which indicate that the value of $\alpha \to 0$ is optimal for the ensemble of SPARA codes. For the iterative decoder, by setting $\alpha \to 0$ the decoding process cannot start; this is the reason why a different value of α yields the optimal result in this case. Disregarding the slight loss of rate due to the high-rate code, the decoder of the SPARA ensemble with $\alpha = \frac{1}{4}$ performs roughly 2 dB away (in terms of $\left(\frac{E_b}{N_0}\right)_2$, when $\left(\frac{E_b}{N_0}\right)_1 = 0$ dB) from its ML decoding threshold (as shown in Fig. 8.4) for a bit error probability of 10^{-3}. This gap between performance under iterative decoding and ML decoding may be bridged by using an *irregular* ensemble, rather than the regular ensemble with $p = 3$ and $q = 6$. Density evolution techniques which were applied to ARA codes [32] may be used to optimize the distribution of the irregular repetition and puncturing patterns. This process improves the performance of these ensembles under iterative decoding, both for the case of a single communication channel and for the case of communicating through parallel channels.

8.4 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 explicit equations for k and β_j (see Eqs. (4.19) and (4.20)) 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 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. (6.6)) 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 (2.1) and (2.4)). 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 are 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. 8.2, 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 $(\nu_1, \ldots, \nu_J)$. 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 chapter 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 Chapters 4 and 5, respectively.

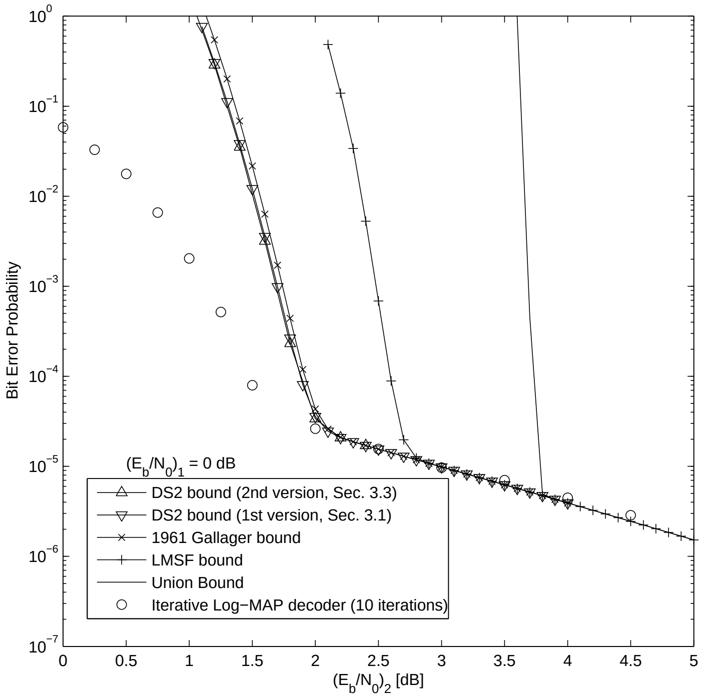


Figure 8.2: Performance bounds for the bit error probability under ML decoding versus computer simulation results of iterative Log-MAP decoding (with 10 iterations). The transmission of this ensemble takes place over two (independent) parallel binary-input AWGN channels. Each bit is equally likely to be assigned to one of these channels, and the energy per bit to spectral noise density of the first channel is set to $\left(\frac{E_\mathrm{b}}{N_0}\right)_1 = 0$ dB. The compared upper bounds on the bit error probability are the generalizations of the DS2 and 1961 Gallager bounds, the LMSF bound from [26], and the union bound (based on (6.6)).

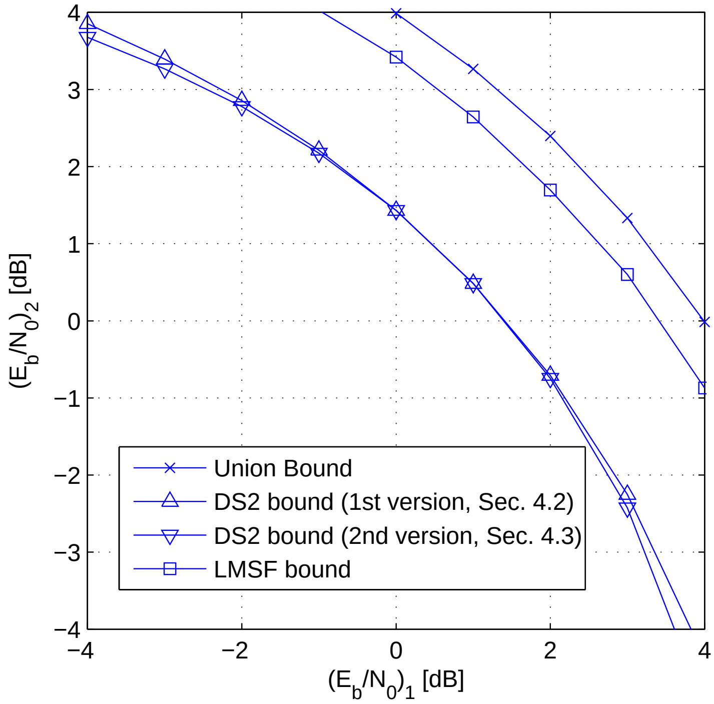


Figure 8.3: Attainable channel regions for the rate one-third uniformly interleaved ensemble of NSRA($N, 3$) codes (see Fig. 2.3 (a)) in the asymptotic case where we let N tend to infinity. The communication takes place over $J = 2$ parallel binary-input AWGN channels, and the bits are equally likely 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 [26], while the boundaries referring to the two versions of the DS2 bound refer to the derivations in Sections 4.2 and 4.3, followed by an optimization of the tilting measures derived in these sections.

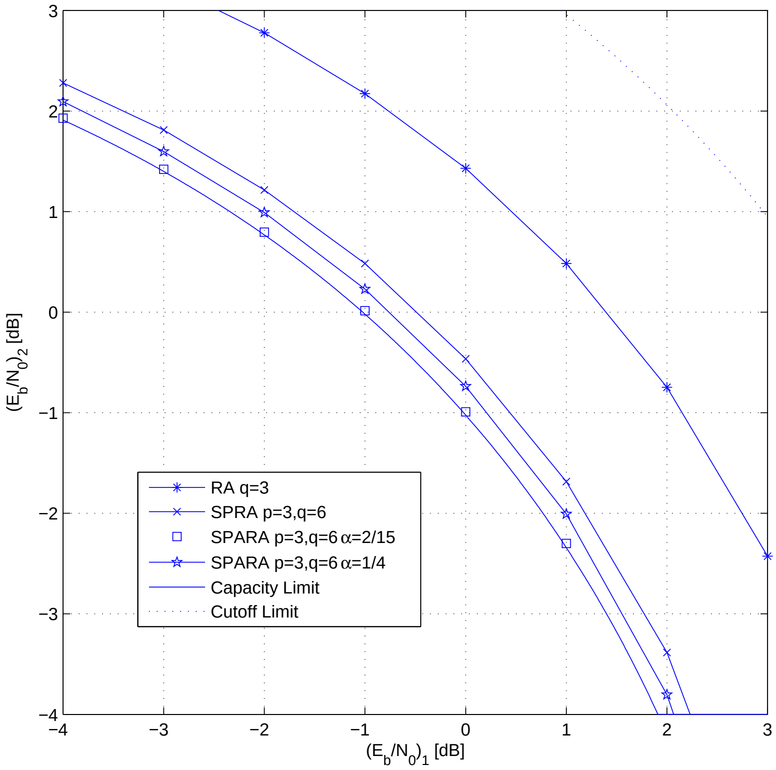


Figure 8.4: Attainable channel regions for the rate one-third uniformly interleaved accumulate-based ensembles with puncturing depicted in Fig. 2.3. These regions refer to the asymptotic case where we let N tend to infinity. The communication takes place over $J = 2$ parallel binary-input AWGN channels, 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 4.2 and 4.3, 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.

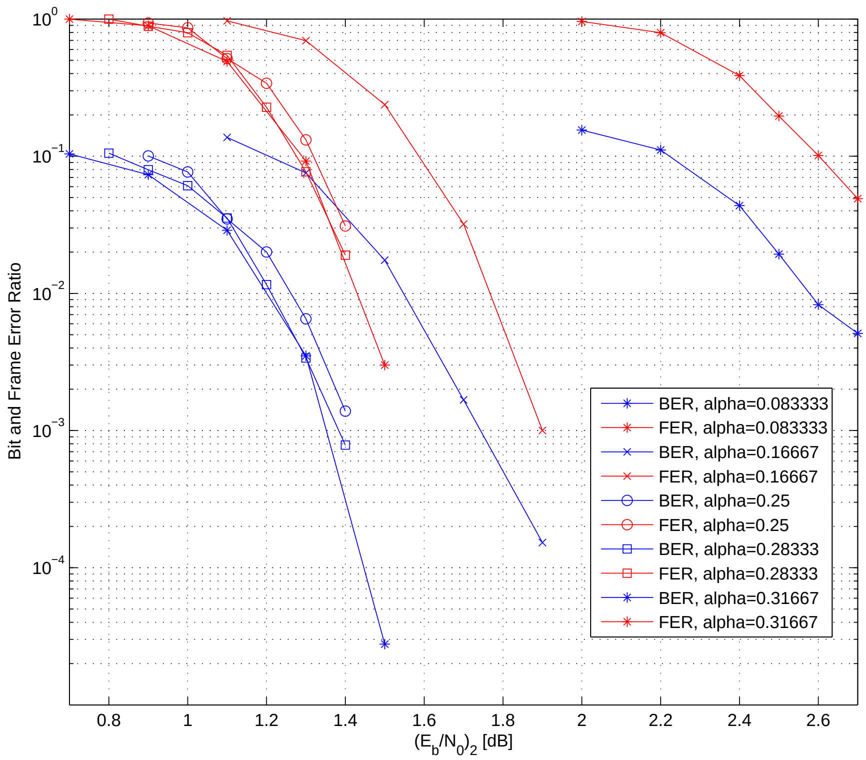


Figure 8.5: Computer simulation results of SPARA codes of blocklength 30000 and rate $\frac{1}{3}$, iteratively decoded using the sum-product algorithm (with 32 iterations). The transmission of this ensemble takes place over two (independent) parallel binary-input AWGN channels. Each bit is equally likely to be assigned to one of these channels, and the energy per bit to spectral noise density of the first channel is set to $\left(\frac{E_\mathrm{b}}{N_0}\right)_1 = 0$ dB.

Chapter 9

Summary and Conclusions

9.1 Contribution of the Thesis

This thesis 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 repeat-accumulate codes and their recent variations which possess low encoding and decoding complexity and exhibit remarkable performance under iterative decoding (see, e.g., [2, 11, 23, 31]). 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 bounding techniques, 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. 4.3) attains the random coding exponent while the other (derived in Sec. 4.2) does not. This difference is attributed to the additional Jensen's inequality in the transition from (4.4) to (4.5) (see p. 41) which is circumvented in the derivation of Sec. 4.3. 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. [26], 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 [26]). The connection between the generalized DS2 bound and the 1961 Gallager bound, which was originally addressed by Divsalar [13] and by Sason and Shamai

[37, 40] for a single channel, is revisited for an arbitrary number of independent parallel channels. In this respect, it is shown that the 1961 Gallager bound [26] is a special case of the generalized DS2 bound derived in Sec. 4.3 and is not a special case of the DS2 bound derived in Sec. 4.2. 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. 4.2 are slightly looser than those computed by the DS2 bound from Sec. 4.3 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. 4.2 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. 8.2 in p. 93).

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. 2.3). The distance spectral analysis serves to assess the performance of these codes under ML decoding where we rely on the bounding techniques developed in this paper and [26] 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 4.2 and 4.3) in Fig. 8.4. This improvement is attributed to the distance spectral thinning effect [30] which is exemplified in Fig. 2.5 by comparing the asymptotic growth rates of the distance spectra for the ensembles in Fig. 2.3 (a)–(c). We also report that for the SPARA ensemble, there is a gap between the ML decoding bounds and computer simulation results under iterative decoding. We believe that this gap stems from the use of an ensemble with regular repetition and puncturing patterns. As observed in [32] for the binary erasure channel, we believe that better results can be achieved by properly selecting irregular repetition and puncturing patterns.

The generalization of the DS2 bound for parallel channels enables to re-derive specific bounds which were originally derived by Liu et al. [26] 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. However, in some cases, the new bounds under ML decoding are a bit pessimistic as compared to computer simulations of sub-optimal iterative decoding (see, e.g., Fig. 8.2), thus indicating that there is still room for further improvement.

The results in this research work are also presented in [34], which was recently accepted for publication in the *IEEE Trans. on Information Theory* (as a full paper).

9.2 Topics for Further Research

In what follows, we point out some possible directions for future research:

- In [6], Bennatan and Burshtein generalized the Shulman and Feder bound to an arbitrary discrete memoryless channel (DMC). They also combine this bound with the union-Bhattacharrya bound, a technique which we use in Chapter 6. A possible direction of research is to generalize the improved bounds, i.e., the DS2 bound and the 1961 Gallager bound to the case of an arbitrary DMC. This generalization may be studied for a single DMC or for the case of parallel DMCs. In the latter case, a random channel mapper can be assumed in order to simplify the analysis.

- Pfister and Sason [32] have recently examined the performance of some ensembles of accumulate-repeat-accumulate (ARA) codes transmitted over the BEC and have obtained results which allow to approach capacity on this channel with *bounded* decoding complexity per information bit. With these results in mind, three directions of research are thus proposed.

  - First, the results in [32] may be generalized to a set of parallel BECs. In this respect, using the technique of a random channel mapper is expected to simplify the analysis, in the same way as we have seen in Chapters 4 and 5. Using the random mapper approach, the density evolution equations should be rewritten so as to accommodate the parallel channel.

- Second, the ML analysis we performed for regular SPARA codes can be extended to irregular ensembles, i.e, ensembles where irregular repetition and puncturing patterns are used. The calculation of bounds under ML decoding requires to extend the calculation of the average distance spectra for irregular ensembles of ARA codes. This will provide a better understanding of the effect of the degree distributions on the gap to capacity.
- Finally, and in continuation to the last direction of research, we have demonstrated that for the ensemble of SPARA codes discussed in Section 8.3, there is a considerable gap between the (upper bound on) performance under ML decoding and practical performance under iterative decoding. We believe this gap stems from the use of regular repetition and puncturing patterns (as is the case for regular LDPC ensembles where the gap between thresholds under ML and iterative decoding is rather large). Allowing these patterns to be *irregular*, as in [32], may enable to bridge this gap. Optimized repetition and puncturing degree distributions may be obtained using density evolution techniques, and the performance could be compared with that of ML decoding.

Appendix 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 (4.18) and demonstrate that they are not even functions. From (4.8), we get the expression

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

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

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

Substituting (4.18) 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 (5.21) in the LHS of (A.2) and manipulate the expression to get

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

Clearly, this expression does not constitute an even function.

Appendix B

Technical Details for Calculus of Variations on (5.26)

The bound on the decoding error probability for constant Hamming weight codes is given by substituting (5.25) into (5.14). Disregarding the multiplicative term $2^{h(\rho)}$, we minimize the expression

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

Employing calculus of variations, we substitute in (B.1) 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

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

Defining the constants

$$
\begin{aligned}
c_1 &\triangleq A_h e^{-nrd} hr \left[\sum_{j=1}^{J} \alpha_j \sum_y \big(p(y|0;j) p(y|1;j) \big)^{\frac{1-r}{2}} f_0(y;j)^r \right]^{h-1} \\
c_2 &\triangleq \left[\sum_{j=1}^{J} \frac{\alpha_j}{2} \sum_y \left(p(y|0;j)^{1-r} + p(y|1;j)^{1-r} \right) f_0(y;j)^r \right]^{n-h} \\
c_3 &\triangleq A_h e^{-nrd} \frac{r(n-h)}{2} \left[\sum_{j=1}^{J} \frac{\alpha_j}{2} \sum_y \left(p(y|0;j)^{1-r} + p(y|1;j)^{1-r} \right) f_0(y;j)^r \right]^{n-h-1} \\
c_4 &\triangleq \left[\sum_{j=1}^{J} \alpha_j \sum_y \big(p(y|0;j) p(y|1;j) \big)^{\frac{1-r}{2}} f_0(y;j)^r \right]^{h} \\
c_5 &\triangleq e^{-nsd} \frac{ns}{2} \left[\sum_{j=1}^{J} \frac{\alpha_j}{2} \sum_y \left(p(y|0;j)^{1-s} + p(y|1;j)^{1-s} \right) f_0(y;j)^s \right]^{n-1}
\end{aligned}
\tag{B.3}
$$

and requiring that the integrand in (B.2) be equal to zero, we get the equivalent condition

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

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 (5.27).

Appendix C

Proof of Theorem 3

The concept of the proof of this theorem is similar to the proof introduced in [25, pp. 40–42] for the single channel case, and the proofs of [26, 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 4.2).

We begin by rewriting the DS2 bound for a specific constant Hamming-weight subcode (4.11) as

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

where

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

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 $E^{\mathrm{DS2}_1}(\delta)$ by

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

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

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

for any $\alpha > 0$. This follows since for weights up to D_n we set $B_h^{\text{opt}} = 1$, 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 (6.1)). Let us examine the behavior of each of the three terms in (C.3). As we let n tend to infinity, the first term in (C.3) goes to 0 due to the third condition of the theorem.

The second term may be rewritten as

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

By the fourth condition of the theorem, the exponent is bounded above by

$$h\left(\ln(\gamma) + \frac{r^{[\mathcal{C}]}(\delta)}{\delta} + o\left(\frac{D_n}{n}\right)\right).$$

Now, the summation only has terms corresponding to $h > D_n$, so $o\left(\frac{D_n}{n}\right) \leq o(1)$. The second condition implies that for small enough α, the exponent is negative and bounded away from 0, say by $-\theta_0$, where $\theta_0 > 0$. Then

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

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

Finally, by using (C.2), the third term in (C.3) may be expressed as

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

which vanishes as $n \to \infty$ due to (7.1), thus completing the proof. □

References

[1] A. Abbasfar, D. Divsalar and K. Yao, "Accumulate repeat accumulate codes," *Proceedings IEEE 2004 Global Telecommunications Conference (GLOBECOM 2004)*, pp. 509–513, 29 November–3 December, 2004, Dallas, Texas, USA.

[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] A. Bennatan and D. Burshtein, "On the application of LDPC codes to arbitrary discrete-memoryless channels," IEEE Trans. on Information Theory, vol. 50, no. 3, pp. 417-438, March 2004.

[7] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near shannon limit error-correcting coding and decoding: Turbo-codes," *Proceedings IEEE Internat. Conf. Commun. 1993*, Geneva, Switzerland, May 1993, pp. 1064–1070.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] D. Divsalar, S. Dolinar, and F. Pollara, "Iterative Turbo Decoder Analysis Based on Density Evolution", *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 5, pp. 891–907, May 2001.

[13] 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`.

[14] 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.

[15] 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.

[16] 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`.

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

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

[19] 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.

[20] 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.

[21] 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, November 2005.

[22] 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.

[23] 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.

[24] 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.

[25] 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`.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] H. D. Pfister and I. Sason, "Capacity-achieving ensembles of accumulate-repeat-accumulate codes for the erasure channel with bounded complexity," to appear in the *IEEE Trans. on Information Theory*, June 2007.

[33] 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.

[34] I. Sason and I. Goldenberg, "Coding for parallel channels: Gallager bounds for binary linear codes with applications to repeat-accumulate codes and variations," accepted to *IEEE Trans. on Information Theory*, revised February 2007. [Online]. Available: http://www.ee.technion.ac.il/people/sason/Sason_Goldenberg.pdf.

[35] 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`.

[36] 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.

[37] 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.

[38] 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.

[39] 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.

[40] 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.

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

[42] Moshe Twitto, "Tightened Upper Bounds on the ML Decoding Error Probability of Binary Linear Block Codes and Applications", *M.Sc. Dissertation.* Technion, Israel Institute of Technology, Haifa, June 2006.

[43] 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.

[44] 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.

חסמי גלגר לערוצים מקביליים
ויישומיהם לטכניקות קידוד מודרניות

גולדנברג עידן

חסמי גלגר לערוצים מקביליים

ויישומיהם לטכניקות קידוד מודרניות

חיבור על מחקר

לשם מילוי חלקי של הדרישות לקבלת תואר

מגיסטר למדעים

הנדסת חשמל

גולדנברג עידן

הוגש לסנט הטכניון מכון טכנולוגי לישראל

שבט תשס"ז חיפה פברואר 2007

המחקר נעשה בהדרכת דר' יגאל ששון
בפקולטה להנדסת חשמל

הכרת תודה

תודות חמות מגיעות למנחה שלי, דר' יגאל ששון, על הנחייתו הצמודה והמסורה לכל אורך הדרך. ללא עזרתו, לא היתה עבודה זו מגיעה לכדי השלמה.

ברצוני להודות גם לפרופ' שלמה שמאי ולדר' עופר עמרני על שהשתתפו בועדת הבחינה שלי.

לבסוף, אני מודה לאשתי, טל גולדנברג, על תמיכתה והסבלנות שגילתה לכל אורך תקופת השתלמותי. עבודה זו מוקדשת לה.

עבודת מחקר זו נתמכה על ידי מלגה שנתית בידי אינטל חיפה.

לאשתי, טל.

תקציר

מערכות תקשורת מודרניות רבות נדרשות לפעול בתנאים בהם הערוץ איננו קבוע בזמן אלא משתנה. בתנאים מסוימים מצב זה ניתן לדימוי על ידי שימוש בסט של ערוצים מקביליים, כאשר כל אחד מהערוצים הינו קבוע בזמן, וכאשר מילת הקוד המשודרת מפוצלת לתתי־מלים שכל אחת מהן משודרת דרך אחד מהערוצים המקביליים. מודל מסוג כזה המתאר ערוץ המורכב מקבוצת ערוצים מקביליים מתאים לניתוח של מספר מיתארים הנפוצים בעולם התקשורת, כגון ערוצי FADING BLOCK, מערכות OFDM, קודים מנוקבים, ועוד.

עבור רוב מערכות התקשורת המקודדות, לא ניתן לקבל ביטוי מדויק עבור הסתברות השגיאה תחת פענוח סבירות מירבית או פענוח תת־אופטימאלי מעשי. כדי שניתן יהיה בכל זאת להעריך ביצועים של מערכות כאלו, יש צורך בשימוש בחסמים על ביצועיהן של מערכות אלו, או לחילופין, להשתמש בסימולציות מחשב. החיסרון הבולט בשימוש בסימולציות מחשב ככלי להערכת הביצועים של מערכת הוא משך הזמן הניכר הדרוש להערכת הסתברות שגיאה נמוכה, בעוד ששימוש בחסמים מאפשר קבלת תוצאות בפרק זמן משמעותית קצר יותר. יתרה מזאת, ביטוי אנליטי המהווה חסם על הסתברות השגיאה מהווה כלי עזר תיאורטי והנדסי חשוב, ומאפשר לקבל תובנה הנדסית על מידת ההשפעה של פרמטרי המערכת על ביצועיה.

אחד החסמים העליונים על הסתברות השגיאה הנפוצים ביותר הינו חסם האיחוד. חסם זה מצטיין בפשטותו הרבה וניתן ליישם אותו עבור מגוון רחב של ערוצים ומערכות. חסרונו העיקרי הוא שעבור קודים ארוכים, הוא חסר תועלת בקצבים הגבוהים מקצב הקיטעון של הערוץ. במשך עשרות שנים היה חסם זה מספק עבור רוב מערכות התקשורת הקיימות, למרות החיסרון שהזכרנו לעיל, מהסיבה שרוב מערכות התקשורת המקודדת עבדו באופן יעיל רק בתחום הקצבים הנמוכים מקצב הקיטעון של הערוץ. הצגת קודי טורבו ונגזרותיהם לפני כעשור שינתה את פני הדברים. קודים אלו מאפשרים לקבל הסתברות שגיאה נמוכה ביותר בקצבים הקרובים לקיבול הערוץ, וזאת תחת פענוח איטרטיבי מעשי, הניתן למימוש בסיבוכיות

המתקבלת על הדעת. תוך שנים ספורות הפכו קודים אלו לקודים סטנדרטיים במערכות תקשורת מודרניות רבות, כגון שידורי וידאו לוויני, הדור השלישי של התקשורת התאית, ועוד. עובדה זו היוותה תמריץ חזק לפיתוח חסמים הדוקים יותר על הסתברות השגיאה, ובפרט עבור קודים הפועלים בקצב הגבוה מקצב הקיטעון של הערוץ. קודי הטורבו שהזכרנו לעיל מפוענחים באמצעות שיטות מעשיות תת־אופטימאליות. למרות זאת, יש עניין מעשי בפיתוח חסמים עליונים על הסתברות השגיאה של מפענח הסבירות המרבית האופטימאלי, היות והדבר נותן אינדיקציה לגבי היכולת ברת ההשגה האולטימטיבית של המערכת. כאשר דנים בתקשורת מעל ערוץ בודד עבור אורכי בלוק השואפים לאינסוף, ניתוח הסתברות השגיאה מניב ספי פענוח המתייחסים לפענוח סבירות מרבית. כאשר מרחיבים את הדיון לערוצים מקביליים, מתקבלים בהקשר זה תחומי ערוץ רב־ממדיים שבהם ניתן להשיג תקשורת אמינה. באופן כללי, גם בניתוח הקשור לערוצים מקביליים וגם בניתוח הקשור לערוץ בודד, לא ניתן לאפיין את איזורי ההחלטה של המקלט (הקשורים לקוד) בצורה מלאה. מסיבה זו, חסמים ברי־שימוש יהיו כאלו שיתבססו אך ורק על תכונות בסיסיות כגון ספקטרום המרחקים של הקוד, אותו ניתן לקבל אנליטית עבור מגוון רחב של קודים. גם את ספקטרום המרחקים קשה לעיתים לחשב עבור קוד ספציפי. כדי להתגבר על בעיה זו, מגדירים בדרך כלל צביר של קודים שמכיל את הקוד הרצוי, ועבור צביר זה מחשבים את ספקטרום המרחקים הממוצע, אותו ניתן לחשב לעיתים רבות ביתר קלות. כתוצאה מזאת, רצוי שחסם עליון טוב יהיה ניתן לישום גם על הסתברות השגיאה הממוצעת של צביר קודים, ולא רק על הסתברות השגיאה של קוד ספציפי. שני חסמים עליונים מרכזיים הידועים היום הם החסם של גלגר, שפותח במסגרת עבודת דוקטורט שבוצעה עוד בשנות השישים של המאה הקודמת, והגרסה השניה והמוכללת של חסם Duman־Salehi (המכונה בקיצור חסם DS2) שפותחה בעשור האחרון. שני חסמים אלה הינם חסמים הדוקים על הסתברות השגיאה המתאימים לניתוח ביצועים של קודים הפועלים בקצבים שקרובים לקיבול הערוץ. בנוסף, שניהם מבוססים על ספקטרום המרחקים של הקוד וניתנים לישום גם עבור צבירים של קודים. עבודות מחקר שפורסמו בעשור האחרון מצביעות על כך שחסמים רבים נוספים שפורסמו בספרות הינם מקרים פרטיים של שני החסמים שהזכרנו לעיל, ובפרט חסם גלגר הוא בעצמו מקרה פרטי של חסם DS2 עבור המקרה של ערוץ תקשורת יחיד. בעבודת מחקר שפורסמה לאחרונה, Liu et al. הכלילו את חסם גלגר שהיה ידוע עבור ערוץ יחיד, למקרה של תקשורת בערוצים מקביליים. עבודה זו הציגה תוצאות הקשורות לתחומי ערוץ בהם ניתן להשיג תקשורת אמינה. בשל המורכבות הקשורה בחישוב החסם ההדוק ביותר, התוצאות בעבודה של Liu et al. מבוססות

על מקרים פרטיים של חסם זה שהינם פחות הדוקים מהחסם האופטימאלי. מכאן שהמוטיבציה היא כפולה: מצד אחד, להרחיב את תחומי הערוץ בהם ניתן להשיג תקשורת אמינה על ידי שימוש בחסם גלגר האופטימאלי ולא במקרים פרטיים שלו, ומצד שני להכליל את חסם DS2 למקרה של ערוצים מקביליים ובכך לנסות להדק את החסמים הקיימים עוד יותר.

העבודה מתחלקת לשמונה חלקים מרכזיים. בחלק הראשון, אנו דנים בחישוב ספקטרום המרחקים הממוצע למספר צבירים של קודים. צבירים אלה כוללים את קודי הטורבו שהזכרנו לעיל ובנוסף כוללים גם מספר צבירים של קודים מסוג Repeat־Accumulate שהוצעו בשנים האחרונות בשל העובדה שהם משיגים ביצועים מרשימים תחת פענוח איטרטיבי, ובשל כך שקודים אלה ניחנים בסיבוכיות קידוד ופענוח נמוכות. בחלק השני, אנו מציגים את מודל הערוץ המקבילי בו אנו דנים לכל אורך העבודה ומציגים בנוסף חומר רקע הקשור לחסמים בערוץ יחיד. החלק השלישי דן בהכללה של חסם DS2 למקרה של תקשורת בערוצים מקביליים. בהקשר זה אנו מקבלים שתי תוצאות נפרדות הנובעות משתי גישות ניתוח שונות בעיקרן. כאשר משווים בין ההדיקות של שני החסמים הללו, מתקבלת התוצאה שאף אחד מהם איננו הדוק יותר מהשני באופן גורף (קרי, לכל צביר של קודים), אולם רק אחת מהגרסאות משיגה את קיבול הערוץ עבור צביר הקודים האקראיים. החלק הרביעי דן בהכללה של חסם גלגר לערוצים מקביליים. בפרק זה מתקבל חסם גלגר האופטימאלי לערוצים מקביליים. כמו־כן, לאור הקשרים שבין חסם גלגר לבין חסם DS2 הידועים למקרה של ערוץ בודד, אנו דנים בקשרים שבין החסמים האלה בהקשר של ערוצים מקביליים, כאשר ההשוואה היא אל מול שתי ההכללות של חסם DS2. החלק החמישי דן במקרים פרטיים של חסמי גלגר ו־DS2 לערוצים מקביליים. כמו־כן אנו דנים בתחומי ערוץ בהם ניתן להשיג תקשורת אמינה, וזאת במקרה של תקשורת בערוצים מקביליים וכאשר אורך הבלוק שואף לאינסוף. בהמשך, מוצגות תוצאות נומריות הקשורות לחסמים. תוצאות אלה מצביעות על שיפור ניכר בניבוי הביצועים הקשורים לערוצים מקביליים. השיפור בתוצאות מודגם הן לאורך בלוק סופי, והן למקרה האסימפטוטי שבו אורך הבלוק שואף לאינסוף. לבסוף מופיע סיכום לתוצאות העבודה ומוצעים נושאים להמשך מחקר.