Performance Analysis of Linear Codes under Maximum-Likelihood Decoding: A Tutorial

Performance Analysis of Linear Codes under Maximum-Likelihood Decoding: A Tutorial

Igal Sason

Shlomo Shamai

Department of Electrical Engineering
Technion – Israel Institute of Technology
Haifa 32000, Israel
{*sason, sshlomo*}*@ee.technion.ac.il*

now

the essence of knowledge

Boston – Delft

Foundations and Trends® in Communications and Information Theory

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

A Cataloging-in-Publication record is available from the Library of Congress

The preferred citation for this publication is I. Sason and S. Shamai, Performance Analysis of Linear Codes under Maximum-Likelihood Decoding: A Tutorial, Foundation and Trends® in Communications and Information Theory, vol 3, no 1/2, pp 1–225, 2006

Printed on acid-free paper

ISBN: 1-933019-32-8

Foundations and Trends® in
Communications and Information Theory
Volume 3 Issue 1/2, 2006

Editorial Scope

Foundations and Trends® in Communications and Information Theory will publish survey and tutorial articles in the following topics:

- Coded modulation
- Coding theory and practice
- Communication complexity
- Communication system design
- Cryptology and data security
- Data compression
- Data networks
- Demodulation and Equalization
- Denoising
- Detection and estimation
- Information theory and statistics
- Information theory and computer science
- Joint source/channel coding
- Modulation and signal design
- Multiuser detection
- Multiuser information theory
- Optical communication channels
- Pattern recognition and learning
- Quantization
- Quantum information processing
- Rate-distortion theory
- Shannon theory
- Signal processing for communications
- Source coding
- Storage and recording codes
- Speech and Image Compression
- Wireless Communications

Information for Librarians

Foundations and Trends® in Communications and Information Theory, 2006, Volume 3, 4 issues. ISSN paper version 1567-2190.ISSN online version 1567-2328.Also available as a combined paper and online subscription.

Foundations and Trends$^{®}$ in
Communications and Information Theory
Vol. 3, No 1/2 (2006) 1–225
© 2006 I. Sason and S. Shamai
DOI: 10.1561/0100000009

Performance Analysis of Linear Codes under Maximum-Likelihood Decoding: A Tutorial

Igal Sason and Shlomo Shamai

Department of Electrical Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel {*sason, sshlomo*}*@ee.technion.ac.il*

Abstract

This article is focused on the performance evaluation of linear codes under maximum-likelihood (ML) decoding. Though the ML decoding algorithm is prohibitively complex for most practical codes, the analysis of linear codes under ML decoding allows to predict their performance without resorting to computer simulations. In this article, upper and lower bounds on the error probability of linear codes under ML decoding are surveyed and applied to codes and ensembles. For upper bounds, we discuss various bounds where focus is put on Gallager bounding techniques and their relation to a variety of other reported bounds. Within the class of lower bounds, we address de Caen's based bounds and their improvements, and also consider sphere-packing bounds with their recent improvements targeting codes of moderate block lengths.

Contents

1

A Short Overview

Overview: Upper and lower bounds on the error probability of linear codes under maximum-likelihood (ML) decoding are shortly surveyed and applied to ensembles of codes on graphs. For upper bounds, we focus on the Gallager bounding techniques and their relation to a variety of other known bounds. Within the class of lower bounds, we address de Caen's based bounds and their improvements, and sphere-packing bounds with their recent developments targeting codes of moderate block lengths. This serves as an introductory section, and a comprehensive overview is provided in the continuation of this tutorial.

1.1 Introduction

Consider the classical coded communication model of transmitting one of equally likely signals over a communication channel. Since the error performance of coded communication systems rarely admits exact expressions, tight analytical upper and lower bounds serve as a useful theoretical and engineering tool for assessing performance and for gaining insight into the effect of the main system parameters. As specific good codes are hard to identify, the performance of ensembles of codes is usually considered. The Fano [71] and Gallager [82] bounds were introduced as efficient tools to determine the error exponents of the ensemble of random codes, providing informative results up to the ultimate capacity limit. Since the advent of information theory, the search for efficient coding systems has motivated the introduction of efficient bounding techniques tailored to specific codes or some carefully chosen ensembles of codes. A classical example is the adaptation of the Fano upper bounding technique [71] to specific codes, as reported in the seminal dissertation by Gallager [81] (to be referred to as the 1961 Gallager-Fano bound). The incentive for introducing and applying such bounds has strengthened with the introduction of various families of codes defined on graphs which closely approach the channel capacity limit with feasible complexity (e.g., turbo codes [24], repeat-accumulate codes [1, 54], and low-density parity-check (LDPC) codes [124, 156]). Clearly, the desired bounds must not be subject to the union bound limitation, since for codes of large enough block lengths, these ensembles of turbo-like codes perform reliably at rates which are considerably above the cutoff rate (R_0) of the channel (recalling that union bounds for long codes are not informative at the portion of the rate region above R_0, where the performance of these capacity-approaching codes is most appealing). Although maximum-likelihood (ML) decoding is in general prohibitively complex for long codes, the derivation of upper and lower bounds on the ML decoding error probability is of interest, providing an ultimate indication of the system performance. Further, the structure of efficient codes is usually not available, necessitating efficient bounds on performance to rely only on basic features, such as the distance spectrum and the input-output weight enumeration function (IOWEF) of the examined code (for the evaluation of the block and bit error probabilities, respectively, of a specific code or ensemble). These latter features can be found by analytical methods (see e.g., [127]).

In classical treatments, due to the difficulty in the analytic characterization of optimal codes, random codes were introduced ([71], [82], [83]). This is also the case with modern approaches and practical coding techniques, where ensembles of codes defined on graphs lend themselves to analytical treatment, while this is not necessarily the case for specifically chosen codes within these families. A desirable feature is to identify efficient bounding techniques encompassing both specific codes and ensembles.

In Sections 2–4, we present various reported upper bounds on the ML decoding error probability, and exemplify their improved tightness as compared to union bounds. We demonstrate in Sections 3 and 4 the underlying connections that exist between these bounds whose computation is solely based on the distance spectrum or the IOWEFs of the codes. The focus of this presentation is directed towards the application of efficient bounding techniques on ML decoding performance, which are not subject to the deficiencies of the union bounds and therefore provide useful results at rates reasonably higher than the cutoff rate. In Sections 2–4 and references therein, improved upper bounds are applied to block codes and turbo-like codes. In addressing the Gallager bounds and their variations, we focus in [183] (and more extensively in Section 4) on the Duman and Salehi variation which originates from the standard Gallager bound. A large class of efficient recent bounds (or their Chernoff versions) is demonstrated to be a special case of the generalized second version of the Duman and Salehi bounds. Implications and applications of these observations are addressed in Section 4.

In Sections 5 and 6, we address lower bounds on the ML decoding error probability and exemplify these bounds on linear block codes. Here we overview a class of bounds which are based on de Caen's bound and its improved version. We also review classical sphere-packing bounds and recent improvements for finite length codes.

We note that every section is self-contained and consequently, notations may (slightly) change from one section to another.

1.2 General approach for the derivation of improved upper bounds

In Sections 3–4, we present many improved upper bounds on the ML decoding error probability which are tighter than the union bound. The basic concept which is common to the derivation of the upper bounds within the class discussed in Sections 3 and 4 is the following:

$$\begin{aligned}\Pr(\text{error}) &= \Pr(\text{error}, \underline{y} \in \mathcal{R}) + \Pr(\text{error}, \underline{y} \notin \mathcal{R}) \\ &\leq \Pr(\text{error}, \underline{y} \in \mathcal{R}) + \Pr(\underline{y} \notin \mathcal{R}) \qquad (1.1)\end{aligned}$$

where $\underline{y}$ is the received signal vector, and $\mathcal{R}$ is an arbitrary region around the transmitted signal point which is interpreted as the "good region". The idea is to use the union bound only for the joint event where the decoder fails to decode correctly, and in addition, the received signal vector falls inside the region $\mathcal{R}$ (i.e., the union bound is used for upper bounding the first term in the right-hand side (RHS) of (1.1)). On the other hand, the second term in the RHS of (1.1) represents the probability of the event where the received signal vector falls outside the region $\mathcal{R}$. This term which is typically the dominant one for low SNR, is not part of the event for which the union bound is used. We note that in the case where the region $\mathcal{R}$ is the whole observation space, the basic approach which is suggested above particularized to the union bound. However, since the upper bound in (1.1) is valid for an arbitrary region $\mathcal{R}$ in the observation space, various improved upper bounds can be derived by an appropriate selection of this region. These bounds could be therefore interpreted as geometric bounds (see [50] and [183]). As we will see, the choice of the region $\mathcal{R}$ is very significant in this bounding technique; different choices of this region have resulted in various different improved upper bounds which are considered extensively in Sections 3 and 4. For instance, the tangential bound of Berlekamp [22] used the basic inequality in (1.1) to provide a considerably tighter bound than the union bound at low SNR values. This was achieved by determining the region $\mathcal{R}$ as a boundary of a plane. For the derivation of the sphere bound [90], Herzberg and Poltyrev have chosen the region $\mathcal{R}$ in (1.1) to be a sphere centered at the transmitted signal vector, and optimized the radius of the sphere in order to get the tightest upper bound within this form. The bound of Divsalar [50] is another simple and tight bound which relies on the basic inequality (1.1). The geometrical region $\mathcal{R}$ in his bound was chosen to be a sphere whose center does not necessarily coincide with the transmitted signal vector, so its radius and the location of its center are jointly optimized in order to provide the tightest bound within this form. Finally, the tangential-sphere bound (TSB) which was proposed for binary linear block codes [152] and for M-ary PSK block coded-modulation schemes [91] selected $\mathcal{R}$ as a circular cone whose central line passes through the origin and the transmitted signal vector. It is one of the tightest upper bounds known to-date for linear codes which are modulated by equi-energy signals and whose transmission takes place over a binary-input AWGN channel (see Fig. 1.1 and [168, 170, 223]).

We note that the bounds mentioned above are only a sample of various bounds reported in Section 3; all of these bounds rely on the inequality (1.1) where the geometric region $\mathcal{R}$ characterizes the resulting upper bounds on the decoding error probability. After providing the general approach, we outline some connections between these bounds and demonstrate a few possible applications.

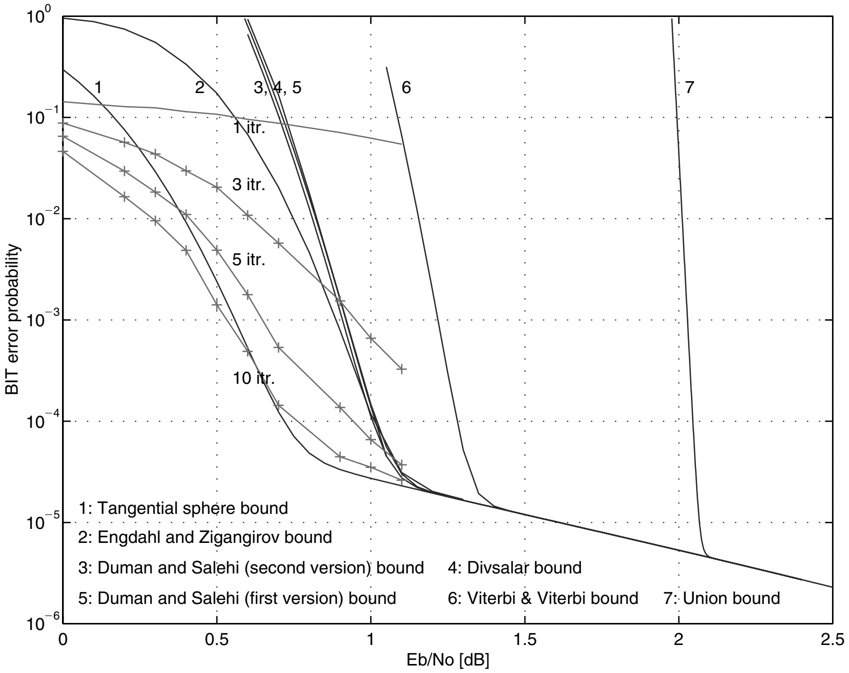

Fig. 1.1 Various bounds for the ensemble of rate$-\frac{1}{3}$ turbo codes whose components are recursive systematic convolutional codes with generators $G_1(D) = G_2(D) = \left[1, \frac{1+D^4}{1+D+D^2+D^3+D^4}\right]$. There is no puncturing of the parity bits, and the uniform interleaver between the two parallel concatenated (component) codes is of length 1000. It is assumed that the transmission of the codes takes place over a binary-input AWGN channel. The upper bounds on the bit error probability under optimal ML decoding are compared with computer simulations of the iterative Log-MAP decoding algorithm with up to 10 iterations.

1.3 On Gallager bounds: Variations and applications

In addressing the Gallager bounding techniques and their variations, we focus in Section 4 on variations of the Gallager bounds and their applications.

In the following, we present shortly the 1965 Gallager bound [82]. Suppose an arbitrary codeword $\underline{x}^m$ (of length-N) is transmitted over a channel. Let $\underline{y}$ designate the observation vector (of N components), and $p_N(\underline{y}|\underline{x}^m)$ be the channel transition probability measure. Then, the conditional ML decoding error probability is given by

$$P_{\mathrm{e}|m} = \sum_{\underline{y}\colon \left\{\exists\, m' \neq m\colon\, p_N(\underline{y}|\underline{x}^{m'}) \geq p_N(\underline{y}|\underline{x}^m)\right\}} p_N(\underline{y}|\underline{x}^m).$$

If the observation vector $\underline{y}$ is such that there exists $m' \neq m$ so that $p_N(\underline{y}|\underline{x}^{m'}) \geq p_N(\underline{y}|\underline{x}^m)$, then for arbitrary $\lambda, \rho \geq 0$, the value of the expression

$$\left(\sum_{m' \neq m} \left(\frac{p_N(\underline{y}|\underline{x}^{m'})}{p_N(\underline{y}|\underline{x}^m)}\right)^{\lambda}\right)^{\rho}$$

is clearly lower bounded by 1, and in general, it is always non-negative. The 1965 Gallager bound [82, 83] therefore states that

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

This upper bound is usually not easily evaluated in terms of basic features of particular codes, except for example, orthogonal codes and the special case of $\rho = 1$ and $\lambda = \frac{1}{2}$ (which yields the Bhattacharyya-union bound).

An alternative bounding technique which originates from the 1965 Gallager bound is the second version of the Duman and Salehi (DS2) bound (see [60, 183]). This bound is calculable in terms of the distance spectrum, not requiring the fine details of the code structure. A similar upper bound on the bit error probability is expressible in terms of the IOWEFs of the codes (or the average IOWEFs of code ensembles). By generalizing the framework of the DS2 bound, a large class of efficient bounds (or their Chernoff versions) is demonstrated to follow from this bound. Implications and applications of these observations are pointed out in [183], including the fully interleaved fading channel, resorting to either matched or mismatched decoding. The proposed approach can be generalized to geometrically uniform non-binary codes, finite state channels, bit-interleaved coded-modulation systems, parallel channels [122], and it can be also used for the derivation of upper bounds on the conditional decoding error probability. In Section 4, we present the suitability of variations on the Gallager bounds as bounding techniques for random and deterministic codes, which partially rely on insightful observations made by Divsalar [50]. Focus is put in [183] on geometric interpretations of the 1961 Gallager-Fano bound (see [71] and [81]). The interconnections between many reported upper bounds are illustrated in Section 4, where it is shown that the generalized DS2 bound particularizes to these upper bounds by proper selections of the tilting measure. Further details, extensions and examples are provided in Section 4.

The TSB [152] happens often to be the tightest reported upper bound for block codes which are transmitted over the binary-input additive white Gaussian noise (AWGN) channel and ML decoded (see e.g., [168] and [170]). However, in the random coding setting, it fails to reproduce the random coding error exponent (see [152]), while the DS2 bound does. In fact, also the Shulman-Feder bound [187] which is a special case of the latter bound achieves capacity for the ensemble of fully random block codes. This substantiates the claim that there is no uniformly best bound. However, we note that the loosened version of the TSB [50] (which involves the Chernoff inequality) maintains the asymptotic (i.e., for infinite block length) exponential tightness of the TSB of Poltyrev [152], and it is a special case of the DS2 bound.

In the following, we exemplify the use of the DS2 bounding technique for fully interleaved fading channels with faulty measurements of the fading samples.

Example 1.1. The Generalized DS2 bound for the Mismatched Regime. In [183], we apply the generalized DS2 bound to study the robustness of a mismatched decoding that is based on ML decoding with respect to the faulty channel measurements. We examine here the robustness of the decoder in case that a BPSK modulated signal is transmitted through a fully interleaved Rayleigh fading channel. For simplicity, the bounds are applied to the case of perfect phase estimation of the i.i.d fading samples (in essence reducing the problem to a real channel). We also assume here that the estimated and real magnitudes of the Rayleigh fading samples have a joint distribution of two correlated bivariate Rayleigh variables with an average power of unity.

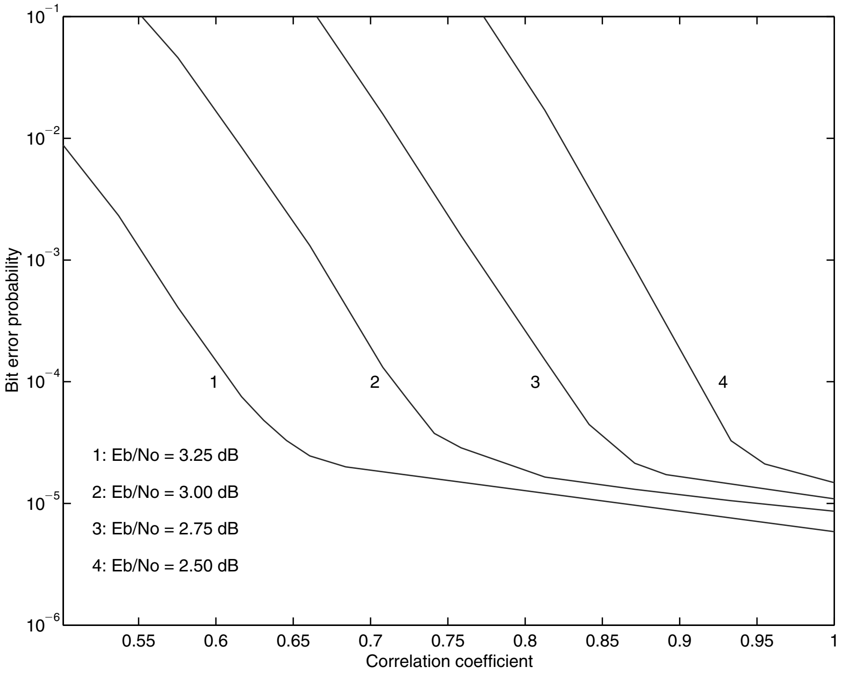

Fig. 1.2 A comparison between upper bounds on the bit error probability for the ensemble of turbo codes considered in Example 1.1 where the transmission of these codes takes place over a fully interleaved Rayleigh fading channel with mismatched decoding. The bounds are based on the combination of the generalized DS2 bound and the tight form of the union bound applied to every constant Hamming-weight subcode. These bounds are plotted for $\frac{E_{\mathrm{b}}}{N_0} = 2.50, 2.75, 3.00$ and 3.25 dB, as a function of the correlation coefficient between the actual i.i.d Rayleigh fading samples and their Rayleigh distributed estimations.

The bounds in Fig. 1.2 refer to the ensemble of uniformly interleaved rate $-\frac{1}{3}$ turbo codes whose components are recursive systematic convolutional codes: $G_1(D) = G_2(D) = \left[1, \frac{1+D^4}{1+D+D^2+D^3+D^4}\right]$ without puncturing of parity bits, and an interleaver length of $N = 1000$. Since for a fully interleaved Rayleigh fading channel with *perfect* side information on the fading samples, the matched channel cutoff rate corresponds to $\frac{E_b}{N_0} = 3.23$ dB then, according to the upper bounds depicted in Fig. 1.2, the ensemble performance of these turbo codes (associated with the ML decoding) is sufficiently robust in case of mismatched decoding, even in a portion of the rate region exceeding the channel matched cutoff rate. The proposed upper bounds depicted here were efficiently implemented in software, thus indicating their feasible computational complexity.

1.4 Lower bounds on the decoding error probability

1.4.1 De Caen inequality and variations

D. de Caen [42] suggested a lower bound on the probability of a finite union of events. While an elementary result (essentially, the Cauchy-Schwartz inequality), it was used to compute lower bounds on the decoding error probability of linear block codes via their distance distribution (see [108] for the binary symmetric channel (BSC), and [182] for the Gaussian channel). In [39], Cohen and Merhav improved de Caen's inequality by introducing an arbitrary non-negative weighting function which is subject to optimization. The concept of this improved bound is presented in the following statement and, like de Caen's inequality, it follows from the Cauchy-Schwartz inequality.

Theorem 1.2. [39, Theorem 2.1] Let $\{A_i\}_{i \in \mathcal{I}}$ be an arbitrary set of events in a probability space $(\Omega, \mathcal{F}, P)$, then the probability of the union of these events is lower bounded by

$$P\left(\bigcup_{i \in \mathcal{I}} A_i\right) \geq \sum_{i \in \mathcal{I}} \left\{ \frac{\left(\displaystyle\sum_{x \in A_i} p(x) m_i(x)\right)^2}{\displaystyle\sum_{j \in \mathcal{I}} \sum_{x \in A_i \cap A_j} p(x) m_i(x)^2} \right\}$$

where m_i is an arbitrary non-negative function on Ω such that the sums in the RHS converge. Further, equality is achieved when

$$m_i(x) = m^*(x) \triangleq \frac{1}{\deg(x)}\,, \quad \forall\, i \in \mathcal{I}$$

where for each $x \in \Omega$

$$\deg(x) \triangleq |\{i \in \mathcal{I} \mid x \in A_i\}|.$$

The lower bound on the union of events in Theorem 1.2 particularizes to de Caen's inequality by the particular choice of the weighting functions $m_i(x) = 1$ for all $i \in \mathcal{I}$, which then gives

$$P\left(\bigcup_{i \in \mathcal{I}} A_i\right) \geq \sum_{i \in \mathcal{I}} \frac{P(A_i)^2}{\sum_{j \in \mathcal{I}} P(A_i \cap A_j)}\,.$$

Cohen and Merhav relied on Theorem 1.2 for the derivation of improved lower bounds on the decoding error probability of linear codes under optimal ML decoding. They exemplified their bounds for BPSK modulated signals which are equally likely to be transmitted among M signals, and the examined communication channels were a BSC and an AWGN channel. In this context, the element x in Theorem 1.2 is replaced by the received vector $\underline{y}$ at the output of the communication channel, and A_i (where $i = 1, 2, \ldots, M - 1$) consists of all the vectors which are closer in the Euclidean sense to the signal $\underline{s}^i$ rather than the transmitted signal $\underline{s}^0$. Following [182], the bounds in [39] get (after some loosening in their tightness) final forms which solely depend on the distance spectrum of the code. Recently, two lower bounds on the ML decoding error probability of linear binary block codes were derived by Behnamfar et al. [16] for BPSK-modulated AWGN channels. These bounds are easier for numerical calculation, but are looser than Cohen-Merhav bounds for low to moderate SNRs.

Note that de Caen's based lower bounds on the decoding error probability (see [16], [39], [108] and [182]) are applicable for *specific* codes but not for ensembles; this restriction is due to the fact that Jensen's inequality does not allow to replace the distance spectrum of a linear code in these bounds by the average distance spectrum of ensembles.

1.4.2 Sphere-packing bounds revisited for moderate block lengths

In the asymptotic case where the block length of a code tends to infinity, the best known lower bound on the decoding error probability for discrete memoryless channels (DMCs) with high levels of noise is the 1967 sphere-packing (SP67) bound [184]. Like the random coding bound of Gallager [82], the sphere-packing bound decreases exponentially with the block length. Further, the error exponent of the SP67 bound is a convex function of the rate which is known to be tight at the portion of the rate region between the critical rate (R_c) and the channel capacity; for this important rate region, the error exponent of the SP67 bound coincides with the error exponent of the random coding bound [184, Part 1]. For the AWGN channel, the 1959 sphere-packing (SP59) bound was derived by Shannon [185] by showing that the error probability of any code whose codewords lie on a sphere must be greater than the error probability of a code of the same length and rate whose codewords are uniformly distributed over that sphere.

The reason that the SP67 bound fails to provide useful results for codes of small to moderate block length is due to the original focus in [184] on asymptotic analysis. In their paper [204], Valembois and Fossorier have recently revisited the SP67 bound in order to make it applicable for codes of moderate block lengths, and also to extend its field of application to continuous output channels (e.g., the AWGN channel which is the communication channel model of the SP59 bound of Shannon [185]). The motivation for the study in [204] was strengthened due to the outstanding performance of codes defined on graphs with moderate block length. The remarkable improvement in the tightness of the SP67 bound was exemplified in [204] for the case of the AWGN channel with BPSK signaling, and it was shown that in some cases, the improved version of the SP67 bound presents an interesting alternative to the SP59 bound [185].

2

Union Bounds: How Tight Can They Be?

Overview: Union bounds are shortly reviewed in this section, and their limited tightness for ensembles of turbo-like codes is addressed.

2.1 Union bounds

Union bounds are based on the trivial inequality which states that the probability of a union of events is upper bounded by the sum of the probabilities of the individual events. Let $\{A_i\}_{i=1}^{M}$ designate a set of M events, then we get the inequality

$$\Pr\left(\bigcup_{i=1}^{M} A_i\right) \leq \sum_{i=1}^{M} \Pr(A_i). \tag{2.1}$$

It is evident that (2.1) turns to be an equality if these events are disjoint. Otherwise, it could be a very loose bound on the probability of a union of events (for instance, the RHS of (2.1) may exceed unity which makes the union bound useless).

Let us consider a binary linear code $\mathcal{C}$ whose transmission takes place over a memoryless, binary-input and output-symmetric (MBIOS) channel. Since the linearity of the code and the symmetry of the channel imply that the error probability under ML decoding does not depend on the transmitted codeword, then the average decoding error probability is equal to the conditional error probability given that the all-zero codeword was transmitted. Let $\{\underline{c}^0, \underline{c}^1, \ldots, \underline{c}^{M-1}\}$ be the set of codewords of the code $\mathcal{C}$ where M is the number of codewords, and let $\underline{c}^0$ be the all-zero codeword which without loss of generality is assumed to be the transmitted codeword. Let $\Pr(\underline{c}^0 \to \underline{c}^i)$ be the pairwise error probability, i.e., the probability for deciding that another codeword $\underline{c}^i$ of the code $\mathcal{C}$ (where $1 \le i \le M-1$) is more likely to be the transmitted codeword than the codeword $\underline{c}^0$ which was actually transmitted (the decision is based on the received signal vector at the output of the channel demodulator). Then it follows immediately from the union bound in (2.1) that

$$\Pr(\text{error}) \le \sum_{i=1}^{M-1} \Pr(\underline{c}^0 \to \underline{c}^i). \tag{2.2}$$

The looseness of the union bound stems from the fact that intersections of half-spaces related to codewords other than the transmitted one, are counted more than once. For a discussion on the asymptotic accuracy of the union bound, the reader is referred to [13]. A possible expurgation of the union bound for linear block codes is based on eliminating codewords which are not neighbors to the transmitted codeword (see [2, 3, 6, 25]). This issue is considered in Section 3.2.9

For simplicity, we consider here the case where the codewords are BPSK modulated before their transmission through the channel. The *input-output weight distribution* of an (N,K) binary linear block code $\mathcal{C}$, designated by $A^{\mathcal{C}}_{w,h}$, is the number of codewords which are encoded by K information bits whose (input) Hamming weight is equal to w, and the (output) Hamming weight of the N-bit codeword is h (where $0 \le w \le K$ and $0 \le h \le N$). The *weight distribution* of the code $\mathcal{C}$, call it $S^{\mathcal{C}}_h$, denotes the number of codewords of the code $\mathcal{C}$ which have a Hamming weight h (the weight distribution is also called the *distance spectrum* of the code $\mathcal{C}$, and clearly $S^{\mathcal{C}}_h = \sum_{w=0}^{K} A^{\mathcal{C}}_{w,h}$). Based on the Bhattacharyya bound, the pairwise error probability between any two codewords differing in h positions is upper bounded by z^h where z

stands for the *Bhattacharyya constant*, i.e.,

$$z \triangleq \int_{-\infty}^{\infty} \sqrt{p(y|X=1)\, p(y|X=0)}\, dy \tag{2.3}$$

and $p(y|x)$ is the *pdf* of the MBIOS channel. From (2.2) and the Bhattacharyya bound, union bounds on the block error probability (P_{e}) and the bit error probability (P_{b}) of the code $\mathcal{C}$ get the form

$$P_{\mathrm{e}} \leq \sum_{h=1}^{N} \sum_{w=1}^{K} A_{w,h}^{\mathcal{C}} z^h \ , \tag{2.4}$$

$$P_{\mathrm{b}} \leq \sum_{h=1}^{N} \sum_{w=1}^{K} \frac{w}{K} A_{w,h}^{\mathcal{C}} z^h \ . \tag{2.5}$$

In order to write the union bound in a compact form, let us define the two-variable polynomial

$$A^{\mathcal{C}}(W,Z) \triangleq \sum_{h=0}^{N} \sum_{w=0}^{K} A_{w,h}^{\mathcal{C}} W^w Z^h \tag{2.6}$$

which designates the *input-output weight enumerator function* (IOWEF) of $\mathcal{C}$.[1] Then, the union bounds on the block error probability (2.4) and the bit error probability (2.5) are expressed as

$$P_{\mathrm{e}} \leq A^{\mathcal{C}}(1,z) - 1 \ , \qquad P_{\mathrm{b}} \leq \frac{1}{K} \left. \frac{\partial A^{\mathcal{C}}(W,Z)}{\partial W} \right|_{W=1,\ Z=z} \tag{2.7}$$

where z designates the Bhattacharyya constant in (2.3). The substraction of 1 in the upper bound on the block error probability follows from the fact that the all-zero codeword contributes an addition of one to the IOWEF of the linear code $\mathcal{C}$.

For the binary-input AWGN channel, the Bhattacharyya constant is equal to $z = e^{-\frac{E_{\mathrm{s}}}{N_0}}$ where $\frac{E_{\mathrm{s}}}{N_0}$ stands for the energy per symbol to the one-sided spectral noise density. Clearly, $\frac{E_{\mathrm{s}}}{N_0} = \frac{RE_{\mathrm{b}}}{N_0}$ where $R = \frac{K}{N}$ is the rate of the code $\mathcal{C}$, and $\frac{E_{\mathrm{b}}}{N_0}$ stands for the energy per information bit

[1] A general technique for the calculation of IOWEFs of convolutional codes was proposed by McEliece [127]. For analytical methods to calculate the weight distribution of binary linear block codes, we refer the reader to [38, 45, 46, 112, 190, 216] and references therein.

to the one-sided spectral noise density. The pairwise error probability between two BPSK modulated codewords differing in h positions and coherently detected over the AWGN channel is equal to $Q\left(\sqrt{\frac{2hRE_\mathrm{b}}{N_0}}\right)$ where

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{t^2}{2}}\, dt \tag{2.8}$$

is the probability that a random Gaussian variable with zero mean and unit variance exceeds the value x. Since $Q(x) \le \frac{1}{2}\, e^{-\frac{x^2}{2}}$ for $x \ge 0$, then it follows that for the AWGN channel, multiplying the Bhattacharyya bound in (2.7) by a factor of one-half still gives a valid exponential upper bound. A tighter upper bound on the bit error probability is derived by using Craig's identity [40, 188]

$$Q(x) = \frac{1}{\pi} \int_0^{\frac{\pi}{2}} e^{-\frac{x^2}{2\sin^2\theta}}\, d\theta\,, \quad x \ge 0\,. \tag{2.9}$$

With the aid of (2.9), we obtain the following upper bounds on the block and bit error probabilities of a binary linear block code $\mathcal{C}$

$$P_\mathrm{e} \le \frac{1}{\pi} \int_0^{\frac{\pi}{2}} A^{\mathcal{C}}(W,Z) - 1 \Big|_{W=1,\ Z=e^{-\frac{RE_\mathrm{b}}{N_0 \sin^2\theta}}} d\theta\,, \tag{2.10}$$

$$P_\mathrm{b} \le \frac{1}{\pi K} \int_0^{\frac{\pi}{2}} \frac{\partial A^{\mathcal{C}}(W,Z)}{\partial W} \Big|_{W=1,\ Z=e^{-\frac{RE_\mathrm{b}}{N_0 \sin^2\theta}}} d\theta \tag{2.11}$$

replacing the upper bounds

$$P_\mathrm{e} \le \frac{1}{2}\left[A^{\mathcal{C}}\left(1, e^{-\frac{RE_\mathrm{b}}{N_0}}\right) - 1\right], \quad P_\mathrm{b} \le \frac{1}{2K} \frac{\partial A^{\mathcal{C}}(W,Z)}{\partial W}\Big|_{W=1,\ Z=e^{-\frac{RE_\mathrm{b}}{N_0}}}$$

which are obviously looser.

As a consequence of the conceptual weakness of union bounds, it is natural to expect that they become useless for linear block codes whose dimension is large. The weakness of union bounds is pronounced at low SNR values (they even exceed unity at low SNR, as is later exemplified in Section 3.2.11). For codes of large enough block lengths, the union bound becomes useless at the rate portion between the cutoff rate and the channel capacity, where the performance of capacity-approaching codes (e.g., turbo codes [24, 23] and LDPC codes [124, 156]) is most appealing (see also the tutorial paper [73] which considers coding techniques for linear Gaussian channels, and the drawback of union bounds). This clearly motivates the need for introducing improved upper bounds on the decoding error probability which are considered in the continuation of this tutorial.

Another theoretical property which is related to the looseness of the union bounds is presented in [74]. Consider the case of BPSK modulation and transmission over a binary-input AWGN channel. The minimal value of $\frac{E_{\mathrm{b}}}{N_0}$ for which the union bound on the ML decoding error probability is dominated by the minimum distance term was evaluated in [74]; this value is referred to as the critical point of the code under ML decoding. For the ensemble of fully random block codes (which are known to achieve the Gilbert-Varshamov distance), this critical point is equal to

$$\frac{1}{R}\ln\left(\frac{1}{h^{-1}(1-R)}-1\right)$$

while the corresponding error probability decreases exponentially with the block length N for values of $\frac{E_{\mathrm{b}}}{N_0}$ exceeding this critical point. In the above expression, h^{-1} stands for the inverse of the binary entropy function to the base 2.

2.2 Union bounds for turbo-like codes

For long enough block codes, union bounds are not informative at rates exceeding the cutoff rate of the channel. The reader is referred to various contributions exploring the ensemble performance of turbo-like codes via the union bounding technique, e.g., union bounds on the ensemble performance of uniformly interleaved parallel and serially concatenated (turbo) codes with fixed component codes are studied in [19, 20, 18, 17, 29, 31, 30, 32, 53, 55, 54, 63, 59, 61, 58, 89, 101, 110, 111, 134, 137, 138, 139, 140, 141, 142, 143, 145, 144, 154, 197, 208, 213, 221]. Union bounds on the ensemble performance of uniformly interleaved turbo-like codes whose components are time-varying recursive systematic convolutional codes are studied in [171, 193], and coding theorems for turbo-like ensembles which rely on union bounds are introduced in [54, 104, 198]. In general, union bounds are not informative at low signal to noise ratios; they only provide some insight about the interleaver gain which is obtained by various ensembles of turbo-like codes. In the sequel, we will present numerical results for union bounds as a benchmark in order to exemplify the improvement in the tightness of various upper bounds over union bounds.

Example 2.1. In order to exemplify the weakness of union bounds, let us consider the ensemble of uniformly interleaved repeat-accumulate (RA) codes [54]. This ensemble is defined as follows: the encoder repeats q times the information block of length N, the bits are then permuted by a uniform interleaver of size qN (i.e., it is a probabilistic interleaver which is chosen uniformly at random over all possible interleavers of this size), and finally, the interleaved bits are encoded by an accumulate code (i.e., a truncated rate-1 recursive convolutional encoder with a transfer function $1/(1+D)$). The encoder of this ensemble is shown in the upper plot of Fig. 2.1.

The ensemble $[\mathcal{RA}_q(N)]$ is defined to be the set of $\frac{(qN)!}{(q!)^N N!}$ different RA codes when considering the different possible permutations of the interleaver.[2] The average input-output weight distribution (i.e., the average number of codewords whose information bits are of Hamming weight w and the Hamming weight of these codewords is h) for the ensemble of uniformly interleaved RA codes $\mathcal{RA}_q(N)$ was originally derived in [54, Section 5], and is given by

$$A_{w,h}^{\mathcal{RA}_q(N)} = \frac{\binom{N}{w}\binom{qN-h}{\lfloor \frac{qw}{2} \rfloor}\binom{h-1}{\lceil \frac{qw}{2} \rceil - 1}}{\binom{qN}{qw}}.$$

Therefore, the average distance spectrum of this ensemble is given by

$$S_h^{\mathcal{RA}_q(N)} = \sum_{w=1}^{\min(N, \lfloor \frac{2h}{q} \rfloor)} \frac{\binom{N}{w}\binom{qN-h}{\lfloor \frac{qw}{2} \rfloor}\binom{h-1}{\lceil \frac{qw}{2} \rceil - 1}}{\binom{qN}{qw}}, \qquad \left\lceil \frac{q}{2} \right\rceil \le h \le qN - \left\lfloor \frac{q}{2} \right\rfloor \tag{2.12}$$

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

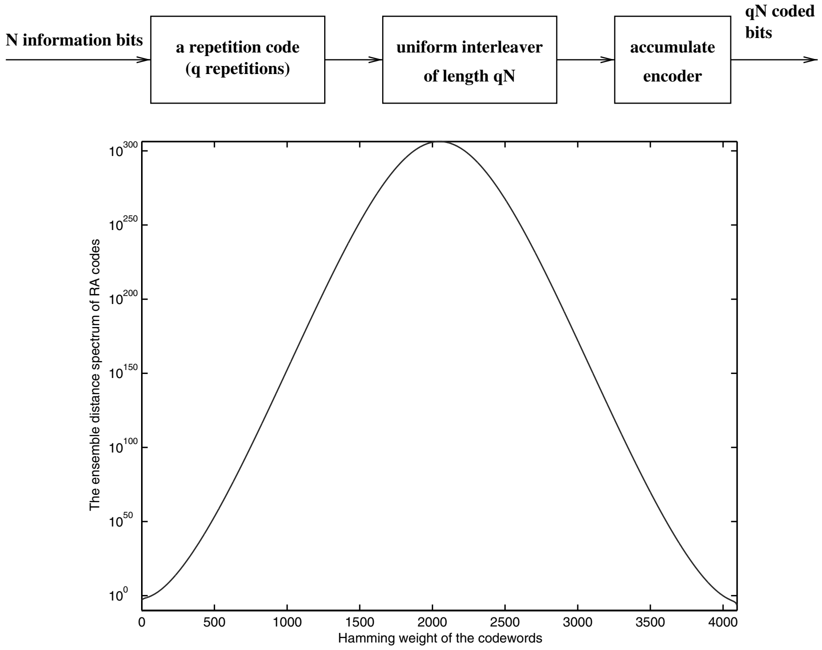

Fig. 2.1 Uniformly Interleaved repeat-accumulate (RA) codes: the upper plot refers to the encoder which forms a serial concatenation of a repetition code with a differential encoder, separated by a uniform interleaver [54]. The lower plot refers to the average distance spectrum of the ensemble, as given in (2.12), for an information block length of $N = 1024$ bits and a number of repetitions of $q = 4$.

and $S_0^{\mathcal{RA}_q(N)} = 1$ since the all-zero vector is always a codeword for any possible choice of the interleaver.

In the following, we compare the union bound with an improved upper bound, the tangential-sphere bound, which is presented in Section 3. This comparison exemplifies the looseness of the union bound at rates exceeding the cutoff rate of the channel. Specifically, for the binary-input AWGN channel, the value of the energy per bit to spectral noise density which corresponds to the cutoff rate is given by

$$\frac{E_{\mathrm{b}}}{N_0} = -\frac{\ln(2^{1-R} - 1)}{R}.$$

For $R = \frac{1}{4}$ bits per channel use (which refers to the code rate of the ensemble depicted in Fig. 2.2), the value of $\frac{E_{\mathrm{b}}}{N_0}$ which corresponds to

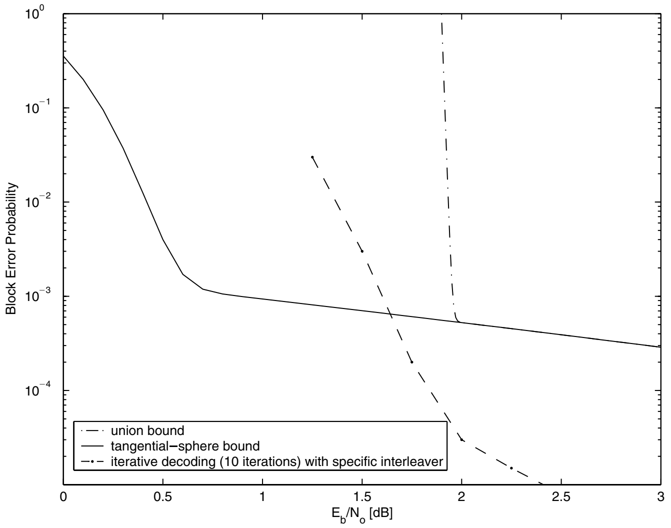

Fig. 2.2 Uniformly Interleaved repeat-accumulate (RA) codes: the union bound is compared with the tangential-sphere bound to be presented in Section 3 and to simulation results under iterative message-passing decoding with 10 iterations. The comparison refers to an RA code with a specific interleaver where the information block length is $N = 1024$ bits and the number of repetitions is $q = 4$ (i.e., the length of the interleaver is $qN = 4096$).

the cutoff rate is equal to 1.85 dB. As is observed in Fig. 2.2, the union bound is useless for values of $\frac{E_{\rm b}}{N_0}$ below 1.85 dB. On the other hand, the tangential-sphere bound which is shown in the same figure demonstrates a remarkable improvement over the union bound for lower values of $\frac{E_{\rm b}}{N_0}$. In order to exemplify the weakness of the union bound as compared to the performance of these codes with a sub-optimal and practical decoder, we combine the compared bounds with simulated results of RA codes under iterative decoding (referring to the sum-product decoding algorithm with 10 iterations). The reason for the improvement of the performance in the error floor region (as compared to the bounds) is due to a specific choice of the interleaver which happens to be better than a uniform interleaver. However, as can be observed from the performance of the iterative decoder, RA codes possess good performance at a certain portion of the rate region between the cutoff rate and the channel capacity (as was first indicated in [54]), where on the other hand, the union bound is useless.

3

Improved Upper Bounds for Gaussian and Fading Channels

Overview: In this section, we present various reported upper bounds on the maximum-likelihood (ML) decoding error probability (including Berlekamp, Divsalar, Duman-Salehi, Engdahl-Zigangirov, Gallager, Hughes, Poltyrev, Sason-Shamai, Shulman-Feder, Viterbi, Yousefi-Khandani and others), and demonstrate the underlying connections that exist between them; the bounds are based on the distance spectra or the input-output weight enumerators of the codes. The focus of this section is directed towards the application of efficient bounding techniques on ML decoding performance, which are not subject to the deficiencies of the union bounds and therefore provide useful results at rates reasonably higher than the cutoff rate, where union bounds are usually useless. We apply improved upper bounds to block codes and turbo-like codes.

3.1 The methodology of the bounding technique

In this section, we present a variety of improved upper bounds which are tighter than the union bound. Let $\underline{y}$ be the received signal vector and let $\mathcal{R}$ be an arbitrary region around the transmitted signal point.

Then the basic concept which is common to all of the improved upper bounds in this section is the following inequality which forms an upper bound on the decoding error probability:

$$\Pr(\text{error}) \leq \Pr(\text{error}, \underline{y} \in \mathcal{R}) + \Pr(\underline{y} \notin \mathcal{R}). \tag{3.1}$$

The region $\mathcal{R}$ in (3.1) is interpreted as the "good region". The idea is to use the union bound on the probability of the joint event where the decoder fails to decode correctly, and in addition, the received signal vector falls inside the region $\mathcal{R}$ (i.e., the union bound is used for upper bounding the first term in the RHS of (3.1)). On the other hand, the second term in the RHS of (3.1) represents the probability of the event where the received signal vector falls outside the region $\mathcal{R}$; this probability is typically the dominant term for very low SNR, and is calculated only one time (since it is not part of the event where the union bound is used). We note that in the case where the region $\mathcal{R}$ is the whole observation space, the basic approach which is suggested above provides the union bound. However, since the upper bound in (3.1) is valid for an arbitrary region $\mathcal{R}$ in the observation space, many improved upper bounds can be derived by an appropriate selection of this region. As we will see, the choice of the region $\mathcal{R}$ is very significant in this bounding technique; different choices of this region have resulted in various different improved upper bounds which are considered in the continuation of this section. For instance, considering the binary-input AWGN channel, the tangential bound of Berlekamp [22] (which is presented in Section 3.2.3) used the basic inequality in (3.1) to provide a considerably tighter bound than the union bound at low SNR values. This was achieved by separating the radial and tangential components of the Gaussian noise with a half-space as the underlying region $\mathcal{R}$. For the derivation of the sphere bound [90], Herzberg and Poltyrev have chosen the region $\mathcal{R}$ in (3.1) to be a sphere around the transmitted signal vector, and optimized the radius of the sphere in order to get the tightest upper bound within this form. The Divsalar bound [50] is another simple and tight bound which relies on the basic inequality (3.1) (the bound is presented in Section 3.2.4). The geometrical region $\mathcal{R}$ in the Divsalar bound was chosen to be a sphere; in addition to the optimization of the radius of this sphere, the center of the sphere which does not necessarily coincide with the transmitted signal vector was optimized as well. Finally, the tangential-sphere bound (TSB) which was proposed for binary linear block codes by Poltyrev [152] and for M-ary PSK block coded-modulation schemes by Herzberg and Poltyrev [91] selected $\mathcal{R}$ as a conical region. It is one of the tightest upper bounds known to-date, and we therefore choose to present it in detail in Section 3.2.1.

We note that the bounds mentioned above are only a sample of the various bounds reported in this section, where all of them rely on the basic inequality (3.1). We have therefore introduced the general approach for the derivation of these bounds before going into the details of the different bounds. In what follows, we present each bound, show connections between these bounds and demonstrate their possible applications.

3.2 Improved upper bounds for the Gaussian channel

3.2.1 The tangential-sphere bound

The TSB was originally derived by Poltyrev [152], and re-derived by Herzberg and Poltyrev who applied the bound to the performance evaluation of PSK block coded modulation schemes [91]. In the following, we present the TSB and discuss its geometrical interpretation (on this occasion, we wish to correct a few printing typos which appear in the derivation of this bound in [91, 152], and enhance the lucidity of its presentation).

Consider a binary linear block code $\mathcal{C}$, and assume that its codewords are mapped to signals with constant energy. Let n and R be the block length and the rate of the code $\mathcal{C}$, respectively, and let E_{b} and E_{s} designate the energy per information bit and per coded symbol, respectively. By assumption, all the transmitted signals can be interpreted as points on an n-dimensional sphere with center at the origin and radius $r_{\mathrm{c}} = \sqrt{nE_{\mathrm{s}}}$ where $E_{\mathrm{s}} = RE_{\mathrm{b}}$. Since the channel is binary-input, output-symmetric and memoryless, and the code $\mathcal{C}$ is binary and linear, then without any loss of generality, one can assume that the all-zero codeword $\underline{c}^0 = (0, 0, \ldots, 0)$ is transmitted (i.e., the conditional error probability does not depend on the transmitted codeword).

We therefore assume in our analysis that $\underline{s}^0$ is the transmitted signal over the AWGN channel.

Referring to Fig. 3.1, let $C_n(\theta)$ designate an n-dimensional circular cone with a half-angle θ whose central line passes through the origin and the transmitted signal ($\underline{s}^0$). Let $\underline{z} = (z_1, z_2, \ldots, z_n)$ designate an n-dimensional noise vector which corresponds to n orthogonal

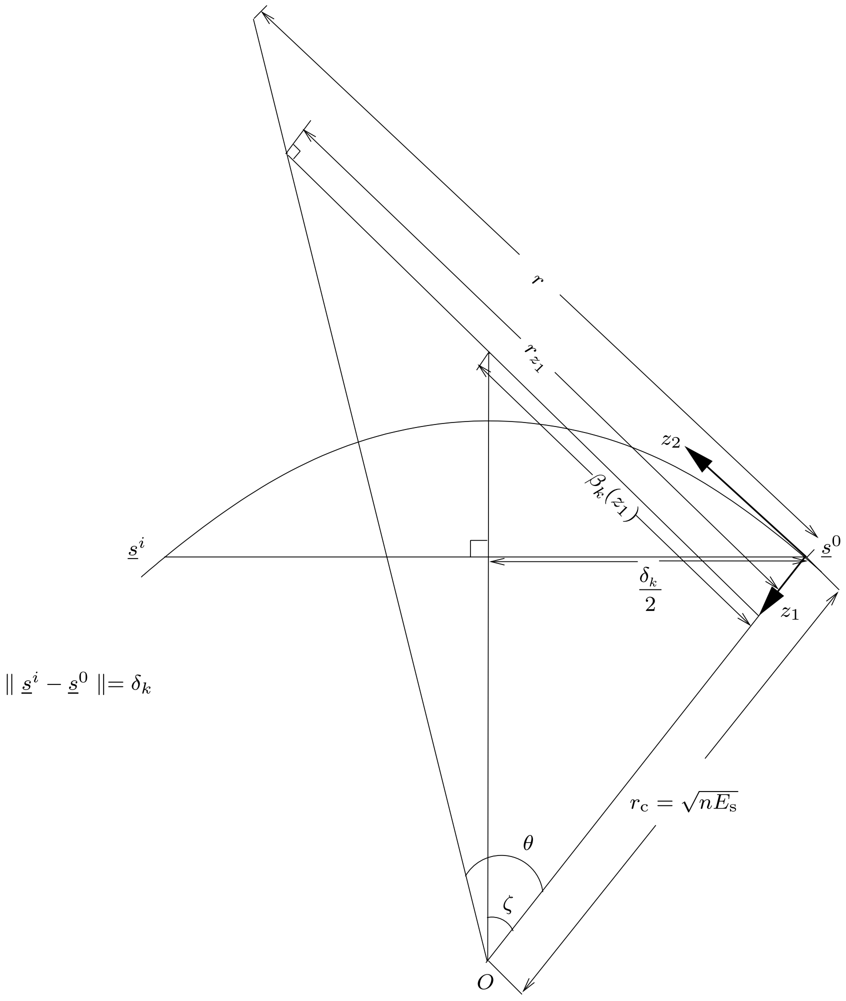

Fig. 3.1 The geometric interpretation of the TSB [91, 152].

projections of the AWGN. Let z_1 be the radial component of $\underline{z}$ (see Fig. 3.1), so the other $n-1$ components of $\underline{z}$ are orthogonal to its radial component. Since $\underline{z}$ is a Gaussian vector and its components are uncorrelated, then the n components of $\underline{z}$ are i.i.d., and each component has a zero mean and variance $\sigma^2 = \frac{N_0}{2}$. From Fig. 3.1, we obtain that

$$\begin{aligned} r &= \sqrt{nE_{\mathrm{s}}}\tan\theta \\ r_{z_1} &= \left(\sqrt{nE_{\mathrm{s}}} - z_1\right)\tan\theta \\ \beta_k(z_1) &= \left(\sqrt{nE_{\mathrm{s}}} - z_1\right)\tan\zeta = \frac{\sqrt{nE_{\mathrm{s}}} - z_1}{\sqrt{nE_{\mathrm{s}} - \frac{\delta_k^2}{4}}}\,\frac{\delta_k}{2}\,. \end{aligned} \tag{3.2}$$

The random variable $Y \triangleq \sum_{i=2}^{n} z_i^2$ is χ^2 distributed with $n-1$ degrees of freedom, so its *pdf* is

$$f_Y(y) = \frac{y^{\frac{n-3}{2}}\, e^{-\frac{y}{2\sigma^2}}\, U(y)}{2^{\frac{n-1}{2}}\sigma^{n-1}\,\Gamma\left(\frac{n-1}{2}\right)}$$

where the function U designates the unit step function, i.e., it is equal to 1 for non-negative arguments, and is zero otherwise. The function Γ designates the complete Gamma function

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt, \quad \mathrm{Real}(x) > 0. \tag{3.3}$$

Conditioned on the value of the radial component of the noise, z_1, let $E(z_1)$ be the event of deciding erroneously (under ML decoding) on a codeword which is different from the transmitted codeword. Given the radial component of the noise, z_1, let $E_k(z_1)$ designate the event of deciding under ML decoding in favor of any other signal ($\underline{s}^i$) whose Euclidean distance from the transmitted signal ($\underline{s}^0$) is equal to δ_k.

Let $\underline{y} = \underline{s}^0 + \underline{z}$ be the received vector at the output of the binary-input AWGN channel. Given the value of the radial component of the noise vector, z_1, the conditional error probability satisfies

$$\Pr\left(E(z_1)|\, z_1\right) \le \Pr\left(E(z_1),\, \underline{y} \in C_n(\theta) \mid z_1\right) + \Pr\left(\underline{y} \notin C_n(\theta) \mid z_1\right) \tag{3.4}$$

and from the union bound

$$\Pr\left(E(z_1),\, \underline{y} \in C_n(\theta) \mid z_1\right) \le \sum_k S_k \Pr\left(E_k(z_1),\, \underline{y} \in C_n(\theta) \mid z_1\right) \tag{3.5}$$

where S_k designates the number of the constant-energy signals $(\underline{s}^i)$ in the considered signal set so that their Euclidean distance from the transmitted signal $(\underline{s}^0)$ is δ_k. We note that for BPSK modulated signals where $\underline{s} = \bigl(2\underline{c} - (1,1,\dots,1)\bigr)\sqrt{E_{\mathrm{s}}}$, the Euclidean distance between the two signals $\underline{s}^i$ and $\underline{s}^0$ is directly linked to the Hamming weight of the codeword $\underline{c}^i$. Let the Hamming distance between the two codewords be k (i.e., $w_{\mathrm{H}}(\underline{c}^i) = k$), then the Euclidean distance between the two BPSK modulated signals is equal to $\delta_k = 2\sqrt{kE_{\mathrm{s}}}$. In the latter case, S_k is the number of codewords of the code $\mathcal{C}$ with Hamming weight k (i.e., $\{S_k\}$ is the distance spectrum of the linear code $\mathcal{C}$).

The combination of Eqs. (3.4) and (3.5) gives

$$\Pr\bigl(E(z_1)|\, z_1\bigr) \le \sum_k \Bigl\{S_k \Pr\bigl(E_k(z_1),\ \underline{y} \in C_n(\theta) \mid z_1\bigr)\Bigr\} + \Pr\bigl(\underline{y} \notin C_n(\theta) \mid z_1\bigr). \qquad (3.6)$$

The second term in the right hand side of (3.6) is easily handled:

$$\begin{aligned}\Pr(\underline{y} \notin C_n(\theta)|\, z_1) &= \Pr(Y > r_{z_1}^2|\, z_1)\\ &= \int_{r_{z_1}^2}^{+\infty} f_Y(y)dy\\ &= \int_{r_{z_1}^2}^{+\infty} \frac{y^{\frac{n-3}{2}} e^{-\frac{y}{2\sigma^2}}}{2^{\frac{n-1}{2}}\, \sigma^{n-1}\, \Gamma\left(\frac{n-1}{2}\right)}\, dy.\end{aligned}$$

This integral is expressible in terms of the incomplete Gamma function

$$\gamma(a,x) \triangleq \frac{1}{\Gamma(a)} \int_0^x t^{a-1} e^{-t}\, dt, \quad a > 0,\ x \ge 0 \qquad (3.7)$$

so we obtain that

$$\Pr\bigl(\underline{y} \notin C_n(\theta)|\, z_1\bigr) = 1 - \gamma\left(\frac{n-1}{2}, \frac{r_{z_1}^2}{2\sigma^2}\right). \qquad (3.8)$$

Let z_2 be the tangential component of the noise vector $\underline{z}$ which is on the plane defined by the two signals $\underline{s}^0$, $\underline{s}^i$ and the origin, and where z_2 is orthogonal to the radial component z_1 (see Fig. 3.1). Referring to the first term in the right hand side of (3.6), it follows from the geometry in Fig. 3.1 that

$$\begin{aligned}\Pr\left(E_k(z_1),\, \underline{y} \in C_n(\theta) \mid z_1\right) &= \Pr\left(E_k(z_1), Y \le r_{z_1}^2 \mid z_1\right) \\ &= \Pr\left(\beta_k(z_1) \le z_2 \le r_{z_1}, Y \le r_{z_1}^2 \mid z_1\right).\end{aligned}$$

Let $V \triangleq \sum_{i=3}^{n} z_i^2$, then $V = Y - z_2^2$, and

$$\Pr\left(E_k(z_1),\, \underline{y} \in C_n(\theta) \mid z_1\right) = \Pr\left(\beta_k(z_1) \le z_2 \le r_{z_1}, V \le r_{z_1}^2 - z_2^2 \mid z_1\right).$$

The random variable V is χ^2 distributed with $n-2$ degrees of freedom, so its *pdf* is equal to

$$f_V(v) = \frac{v^{\frac{n-4}{2}}\, e^{-\frac{v}{2\sigma^2}}}{2^{\frac{n-2}{2}} \sigma^{n-2}\, \Gamma\left(\frac{n-2}{2}\right)}\,, \quad v \ge 0$$

and since the random variables z_2 and V are statistically independent, then

$$\Pr\left(E_k(z_1),\, \underline{y} \in C_n(\theta) \mid z_1\right) = \int_{\beta_k(z_1)}^{r_{z_1}} \frac{e^{-\frac{z_2^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} \int_0^{r_{z_1}^2 - z_2^2} f_V(v)\, dv\, dz_2. \quad (3.9)$$

The calculation of the statistical expectation of both sides of (3.6) with respect to the radial noise component z_1 gives

$$\begin{aligned}P_{\mathrm{e}} &= E_{z_1}\left[\Pr\left(E(z_1) \mid z_1\right)\right] \\ &\le \sum_{k:\, \beta_k(z_1) < r_{z_1}} \left\{ S_k\, E_{z_1}\Big[\Pr\left(E_k(z_1),\, \underline{y} \in C_n(\theta) \mid z_1\right)\Big]\right\} \\ &\qquad + E_{z_1}\Big[\Pr\left(\underline{y} \notin C_n(\theta) \mid z_1\right)\Big]\end{aligned}$$

where the condition $\beta_k(z_1) < r_{z_1}$ in the above sum follows directly from the condition $\zeta < \theta$ (based on the geometry in Fig. 3.1 on p. 24), and therefore is independent of z_1. From (3.8) and (3.9), we obtain the following upper bound on the block error probability under ML decoding which only depends on the distance spectrum of the code:

$$\begin{aligned}P_{\mathrm{e}} \le \int_{-\infty}^{+\infty} \frac{e^{-\frac{z_1^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} \Bigg\{ \sum_{k:\, \frac{\delta_k}{2} < \alpha_k} \left\{ S_k \int_{\beta_k(z_1)}^{r_{z_1}} \frac{e^{-\frac{z_2^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} \int_0^{r_{z_1}^2 - z_2^2} f_V(v)\, dv\, dz_2 \right\} \\ + 1 - \gamma\left(\frac{n-1}{2}, \frac{r_{z_1}^2}{2\sigma^2}\right) \Bigg\}\, dz_1. \quad (3.10)\end{aligned}$$

The determination of the indices k which are taken into account in the summation in the RHS of (3.10) relies on the equivalence between the inequalities $\beta_k(z_1) < r_{z_1}$ and $\frac{\delta_k}{2} < \alpha_k$ where

$$\alpha_k \triangleq r\sqrt{1 - \frac{\delta_k^2}{4nE_\mathrm{s}}}. \tag{3.11}$$

The upper bound (3.10) on the ML decoding error probability is valid for all positive values of r. Hence, the optimal radius r (in the sense of achieving the tightest upper bound) is determined by setting to zero the partial derivative of the right side in (3.10) with respect to r_{z_1}. After tedious but straightforward algebra, we obtain the following optimization equation[1] for the optimal value of the radius r of the TSB:

$$\begin{cases} \displaystyle\sum_{k\colon \frac{\delta_k}{2} < \alpha_k} S_k \int_0^{\varphi_k} \sin^{n-3}\phi \; d\phi = \frac{\sqrt{\pi}\,\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \\ \varphi_k \triangleq \cos^{-1}\left(\frac{\delta_k}{2\alpha_k}\right) \end{cases} \tag{3.12}$$

where α_k is given in (3.11). The nice property of (3.12) is that the optimized value of r *does not depend on the signal to noise ratio, but only on the distance spectrum of the code.* This enables to calculate the optimal value of r for a specific binary linear block code, and then to calculate the TSB in (3.10) for different values of $\frac{E_\mathrm{b}}{N_0}$. We note that the integrals in the left hand side of the optimization equation (3.12) can be handled by invoking the identities in [86, Eqs. (2.511.2) & (2.511.3) (see p. 159)]. It is evident that the optimized TSB does not exceed 1 (since if we let $\theta \to 0$, then the n-dimensional cone $C_n(\theta)$ tends to an empty set; the second term of the RHS in (3.4) tends therefore to 1, which implies the trivial bound $P_\mathrm{e} \leq 1$). This is in contrast to the union bound which diverges at low values of $\frac{E_\mathrm{b}}{N_0}$. A proof for the existence and uniqueness of a solution r to the optimization equation (3.12) was provided in [170, Appendix B], together with an efficient algorithm to

[1] We note that in [152], the $\sqrt{\pi}$ in the first equation of (3.12) was written by mistake in the denominator instead of the numerator, which yields a sub-optimal choice of r in (3.17), and hence a looser upper bound in the final form (3.10) of the TSB.

solve this equation numerically. Sason and Shamai have adapted the TSB to obtain upper bounds on the *bit error probability* of a binary linear code $\mathcal{C}$ (see [170, Appendix C]). To this end, one replaces the Euclidean distance spectrum $\{S_k\}$ in the RHS of (3.10) by

$$S'_k = \sum_{w=1}^{K} \left(\frac{w}{K}\right) A_{w,k}\ , \quad k = d_{\min}, \ldots, n \tag{3.13}$$

where K designates the dimension of the linear code $\mathcal{C}$, and $A_{w,k}$ designates the number of the codewords which are encoded by information bits whose Hamming weight is w and mapped to signals of Euclidean distance δ_k from the transmitted signal.[2] By this replacement, the right hand side of (3.10) becomes an upper bound on the bit error probability of the code $\mathcal{C}$. Since

$$S_k = \sum_{w=1}^{K} A_{w,k}\ , \quad k = d_{\min}, \ldots, n \tag{3.14}$$

then it immediately follows from (3.13) and (3.14) that $S'_k \le S_k$ for all integer values of k, so the resulting upper bound on the bit error probability is clearly smaller than the upper bound on the block error probability, as could be expected. For further observations on the derivation of the TSB-based upper bound on the bit error probability, we refer the reader to [170, Appendix C].

In the following, we suggest a slightly looser upper bound which is considerably easier to calculate. The inner integral in the RHS of (3.9) can be upper bounded with the aid of the incomplete Gamma function, as follows:

$$\begin{aligned}
&\Pr\big(E_k(z_1),\ \underline{y} \in C_n(\theta) \mid z_1\big) \\
&\quad = \int_{\beta_k(z_1)}^{r_{z_1}} \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{z_2^2}{2\sigma^2}}\, \gamma\left(\frac{n-2}{2}, \frac{r_{z_1}^2 - z_2^2}{2\sigma^2}\right) dz_2
\end{aligned}$$

[2] We note that for BPSK modulated signals, $A_{w,k}$ designates the number of codewords of Hamming weight k which are encoded by a block of information bits of Hamming weight w.

$$\begin{aligned}
&\leq \gamma\left(\frac{n-2}{2}, \frac{r_{z_1}^2 - \beta_k^2(z_1)}{2\sigma^2}\right) \int_{\beta_k(z_1)}^{r_{z_1}} \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{z_2^2}{2\sigma^2}} dz_2 \\
&= \gamma\left(\frac{n-2}{2}, \frac{r_{z_1}^2 - \beta_k^2(z_1)}{2\sigma^2}\right) \left[Q\left(\frac{\beta_k(z_1)}{\sigma}\right) - Q\left(\frac{r_{z_1}}{\sigma}\right)\right]. \qquad (3.15)
\end{aligned}$$

This suggested upper bound circumvents the need to compute numerically the inner integral which appears inside the sum of the RHS of (3.10) (such a numerical computation is required for every integer k for which $\delta_k < \frac{\alpha_k}{2}$ and $S_k \neq 0$). Finally, we obtain from (3.10) and (3.15) that the slightly looser version of the TSB upper bound on the block error probability (P_e) is only based on the Euclidean distance spectrum $\{S_k\}$, and it reads:

$$\begin{aligned}
P_\mathrm{e} \leq \int_{-\infty}^{+\infty} \frac{dz_1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{z_1^2}{2\sigma^2}} &\left\{ \begin{aligned} &1 - \gamma\left(\frac{n-1}{2}, \frac{r_{z_1}^2}{2\sigma^2}\right) \\ &+ \sum_{k\colon \frac{\delta_k}{2} < \alpha_k} S_k \left[Q\left(\frac{\beta_k(z_1)}{\sigma}\right) - Q\left(\frac{r_{z_1}}{\sigma}\right)\right] \\ &\qquad\qquad \cdot\gamma\left(\frac{n-2}{2}, \frac{r_{z_1}^2 - \beta_k^2(z_1)}{2\sigma^2}\right) \end{aligned} \right\} \\
&+ Q\left(\sqrt{\frac{2nRE_\mathrm{b}}{N_0}}\right)
\end{aligned} \qquad (3.16)$$

where

$$\begin{cases} \sigma^2 &= \dfrac{N_0}{2} \\ r_{z_1} &= \left(1 - \dfrac{z_1}{\sqrt{nE_\mathrm{s}}}\right) r \\ \beta_k(z_1) &= \dfrac{r_{z_1}}{\sqrt{1 - \dfrac{\delta_k^2}{4nE_\mathrm{s}}}} \cdot \dfrac{\delta_k}{2r} \end{cases} \qquad (3.17)$$

and α_k is given in (3.11). We note that the need for the last summand in the RHS of (3.16) was explained in [170, Appendix A]; it is due to the possibility that the radial component of the noise vector (z_1) is large enough so that we refer to the second side of the cone (i.e., if $z_1 > \sqrt{nE_\mathrm{s}}$). This term didn't appear in the original derivation of the

TSB in [91, 152], but it has typically a negligible effect on the upper bound (see [170, Appendix A]).

In order to significantly reduce the complexity which is involved with the calculation of the TSB and paying a very minor loss in the tightness of this bound, we refer in the continuation to the form of the TSB in (3.16) and the optimization equation in (3.12).

We emphasize that the TSB is not limited to binary linear codes (see e.g., [66] where the TSB is applied to the performance analysis of Reed-Solomon codes). The single property of the coding scheme which is required for the validity of the bound is the equal-energy property, so that the signals are represented by points on an n-dimensional sphere (see Fig. 3.1 on p. 24). We note that the constellation is not restricted to be geometrically uniform; in this case, if the signal set is still of equal energy, then one can use the proposed bounding technique to evaluate the conditional error probability given a particular transmitted signal, provided that the Euclidean distance spectrum with respect to that signal point is available.

In [223], Yousefi and Khandani generalize the derivation of the TSB. To this end, they choose to consider geometrical regions $\mathcal{R}$ in (3.1) which have an azimuthal symmetry with respect to the radial component of the noise. For equal-energy (sphere) signals, they prove that the optimal geometrical region is the conical region of the TSB, and therefore the generalized TSB is equal to the TSB of Poltyrev. This proves the optimality of the cones as geometrical regions with azimuthal symmetry, and therefore the geometrical interpretation of the TSB justifies its tightness. We note however that by the relaxation of the azimuthal symmetry of the region $\mathcal{R}$, the tightness of the bound in (3.1) can be improved. As an example for this possible improvement, we refer to the Shulman and Feder bound [187] which is considered in the next section; it is shown that the geometrical region $\mathcal{R}$ which is associated with this bound does not possess this azimuthal symmetry. However, it reproduces the random coding error exponent of Gallager [82], and therefore achieves capacity for the ensemble of fully random block codes, in contrast to the TSB.

In many communication systems, data is divided into different importance levels, and unequal error protection is desirable for taking into account their different error sensitivities. In [7], the TSB is applied to obtain performance bounds on the ML decoding error probability when binary linear block codes are transmitted over an AWGN channel and the bits are unequally error protected. These bounds are applied to uniformly interleaved turbo codes under ML decoding, and compared with computer simulation results under iterative decoding; this comparison shows a good match, also at a portion of the rate region between the cutoff rate and the channel capacity.

The TSB happens often to be one of the tightest reported upper bounds for block codes which are transmitted over a binary-input AWGN channel and ML decoded (see e.g., [50, 90, 91, 142, 152, 168, 170, 169, 211, 225]). However, in the random coding setting, it fails to reproduce the random coding exponent (especially for high code rates) while the 1965 Gallager bound [82] (which is introduced in Section 4.2.1) achieves the channel capacity for fully random block codes. For further details on the random coding error exponent of the TSB over the binary-input AWGN channel, we refer the reader to [152, Section 5] and [199] (to be discussed later in Section 3.2.10.1).

3.2.2 Improvements on the tangential-sphere bound

Zangl and Herzog [225] improved the TSB on the bit error probability which was derived earlier by Sason and Shamai [170] (as explained in Section 3.2.1). The improvement in [225] refers to the term of the TSB which corresponds to the case where the noise is large enough so that the received signal vector falls outside the conical region around the transmitted signal vector (see Fig. 3.1). The basic idea used for the derivation of the improved upper bound of Zangl and Herzog is to replace the worst case assumption (which assumes that if the received signal vector falls outside the conical region, then all the decoded information bits are wrong) by a more refined calculation which takes into account the actual fraction of information bits which are decoded incorrectly in the latter case. We note that for the parallel and serial concatenated codes which are exemplified in [225], the improvement achieved by the new bound is rather small.

Yousefi and Khandani [224] derived a new upper bound on the block error probability of binary linear block codes whose transmission takes place over a binary-input AWGN channel, and where they are coherently detected and ML decoded. To this end, they used a Bonferroni-type inequality of the second degree [79, 100] (instead of the union bound) to get an upper bound on the joint probability of decoding error and the event that the received signal vector falls within the corresponding conical region around the transmitted signal vector. The basic idea in [224] relies on the inequality which states that if $\{A_i\}_{i=1}^{M}$ designates a set of M events, and A_i^C designates the complementary of the event A_i, then

$$\begin{aligned}
&\Pr\left(\bigcup_{i=1}^{M} A_i\right) \\
&\quad = \Pr(A_1) + \Pr(A_2 \cap A_1^C) + \ldots + \Pr(A_M \cap A_1^C \ldots \cap A_{M-1}^C) \\
&\quad \le \Pr(A_1) + \sum_{i=2}^{M} \Pr(A_i \cap A_{\hat{i}}^C). \qquad (3.18)
\end{aligned}$$

where the indices $\hat{i} \in \{1,2,\ldots,i-1\}$ are chosen arbitrarily for $2 \le i \le M$. Clearly, inequality (3.18) gives a tighter bound on the probability of a union of events than the union bound in (2.1). The concept of the inequality in (3.18) is applied to the first term in the right hand side of (3.4) (instead of the union bound in (3.5) which gives a looser upper bound on the joint probability of decoding error and the event where the received signal vector is inside the corresponding conical region in Fig. 3.1). Since the resulting upper bound in [224, Eq. (25)] cannot be calculated in terms of the distance spectrum of the code (or ensemble of codes), the upper bound is loosened in a way which makes it dependent solely on the distance spectrum. From the final form of the bound in [224, Eqs. (28), (30)–(33)] which involves the enlargement of the original codebook by all n-tuples of Hamming weight w (the parameter w is optimized numerically in the final form of the bound), it is not clear that the latter bound is uniformly better than the TSB. Yousefi and Khandani claim that their new bound is tighter than the TSB [224], but it seems that proving such a claim is not trivial. We note that for two short BCH codes, it was shown numerically that the

TSB is slightly worse than the new bound in [224], but a general proof for this claim is missing. Moreover, the upper bound in [224] is more complex for calculation than the TSB because of the introduction of the additional parameter (w) in the former bound which is optimized numerically (due to the enlargement of the original codebook by all vectors of Hamming weight w).

Recently, Mehrabian and Yousefi proposed in [131] an improved version of the TSB. The derivation of their proposed bound follows along the lines of the bound derived in [224], and the numerical results provided by the authors shows that the resulting improvement of the new bound as compared to the TSB is very marginal, but in parallel its computational complexity as compared to the TSB is significantly higher. It was recently demonstrated in [199] that these improvements over the TSB do not have any implication on the error exponent. A closed form expression for this error exponent, as well as its comparison with the error exponents which are associated with the union bound and the Gallager bound [82], are exemplified in Section 3.2.10.1.

3.2.3 The tangential bound

The tangential bound was derived by Berlekamp [22, pp. 572–574]. It can be seen as a particular case of the TSB (see Section 3.2.1) where we let the radius r of the cone in Fig. 3.1 (see p. 24) tend to infinity. The derivation of the bound relies on the inequality

$$P_{\mathrm{e}} \leq \Pr(\text{error}, z_1 \leq \gamma_0) + \Pr(z_1 > \gamma_0) \tag{3.19}$$

where z_1 is the radial component of the noise vector (see Fig. 3.1), and γ_0 is an arbitrary constant which is later optimized, so as to obtain the tightest bound within this form. To proceed, a union bound like the one in (3.5) (see p. 25) is invoked in order to obtain an upper bound on the first term in the RHS of (3.19). The second term in the RHS of (3.19) is easy to calculate: since $z_1 \sim N(0, \sigma^2)$, then $\Pr(z_1 > \gamma_0) = Q\left(\frac{\gamma_0}{\sigma}\right)$. Referring to Fig. 3.1, we obtain the following inequality from (3.19):

$$P_{\mathrm{e}} \leq \sum_k \Big\{ S_k \Pr\big(z_1 \leq \gamma_0,\ z_2 \geq \beta_k(z_1)\big) \Big\} + Q\left(\frac{\gamma_0}{\sigma}\right) \tag{3.20}$$

and then from (3.2) (see p. 25) and (3.20), we obtain that

$$P_{\mathrm{e}} \leq \sum_k \left\{ S_k \int_{-\infty}^{\gamma_0} Q\left(\frac{\sqrt{nE_{\mathrm{s}}} - z_1}{\sqrt{nE_{\mathrm{s}} - \frac{\delta_k^2}{4}}} \frac{\delta_k}{2\sigma} \right) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{z_1^2}{2\sigma^2}} dz_1 \right\} + Q\left(\frac{\gamma_0}{\sigma}\right). \tag{3.21}$$

The particular choice of $r \to \infty$ in Fig. 3.1 is sub-optimal for the minimization of the TSB (i.e., this choice does not necessarily lead to the tightest upper bound in (3.16)). It therefore follows that the tangential bound is inherently looser than the TSB (as was proved in [91, Lemma 2]). The optimal value of γ_0 in the RHS of (3.21) is calculated by setting its partial derivative with respect to γ_0 to zero, which gives the following equation:[3]

$$\sum_k \left\{ S_k \, Q\left(\frac{\sqrt{nE_{\mathrm{s}}} - \gamma_0}{\sqrt{nE_{\mathrm{s}} - \frac{\delta_k^2}{4}}} \frac{\delta_k}{2\sigma} \right) \right\} = 1. \tag{3.22}$$

In the following, we derive a slightly looser bound which is easier to calculate than the tangential bound in (3.21). Let

$$u = z_1 \sin\zeta + z_2 \cos\zeta, \quad v = -z_1 \cos\zeta + z_2 \sin\zeta \tag{3.23}$$

be the noise components which are received by rotating z_1 and z_2, respectively, in the clockwise direction by an angle $\xi \triangleq \frac{\pi}{2} - \zeta$; it is evident from Fig. 3.1 that $\sin\zeta = \frac{\delta_k}{2\sqrt{nE_{\mathrm{s}}}}$. The condition for an erroneous pairwise ML decoding in Fig. 3.1 is equivalent to the satisfiability of the condition $u \geq \frac{\delta_k}{2}$. It is therefore easy to verify from (3.23) that if the ML decoder decides in favor of $\underline{s}^i$ (instead of the transmitted signal point $\underline{s}^0$) and $z_1 \leq \gamma_0$, then

$$u \geq \frac{\delta_k}{2}\,, \quad v \geq \frac{\frac{\delta_k^2}{4} - \gamma_0\sqrt{nE_{\mathrm{s}}}}{\sqrt{nE_{\mathrm{s}} - \frac{\delta_k^2}{4}}}. \tag{3.24}$$

Since the random variables u and v in (3.23) are un-correlated and jointly Gaussian distributed with zero mean and variance σ^2, then they

[3] Due to the monotonicity of the Q function, the existence and uniqueness of a solution γ_0 in the implicit equation (3.22) is always insured; the monotonicity property makes the numerical solution of γ_0 in (3.22) an easy task (e.g., by using the bisection method).

are statistically independent. Based on (3.20) and (3.24), we obtain the following upper bound on the ML decoding error probability:

$$P_{\mathrm{e}} \leq \sum_k \left\{ S_k \, Q\left(\frac{\delta_k}{2\sigma}\right) \, Q\left(\frac{\frac{\delta_k^2}{4} - \gamma_0\sqrt{nE_{\mathrm{s}}}}{\sqrt{nE_{\mathrm{s}} - \frac{\delta_k^2}{4}}\;\sigma}\right)\right\} + Q\left(\frac{\gamma_0}{\sigma}\right) \tag{3.25}$$

The bound stated in (3.25), together with the corresponding optimized γ_0 (computed by setting to zero the partial derivative of the function in the RHS of (3.25) with respect to γ_0), form a simplified version of the tangential bound. Although the tangential bound (and hence, its simplified version) are always looser than the TSB, they are both uniformly tighter than the union bound (as was proved in [91, Lemma 1]). This statement is easily proved by letting γ_0 in the simplified version of the tangential bound (3.25) tend to infinity; if $\gamma_0 \to \infty$, then we obtain from (3.25) that $P_{\mathrm{e}} \leq \sum_k S_k \, Q\left(\frac{\delta_k}{2\sigma}\right)$ which coincides with the union bound, and therefore shows that the simplified version of the tangential bound (3.25) (and hence, the tangential bound in (3.21) and (3.22)) is always tighter than the union bound. We note that in the case where $\gamma_0 \to -\infty$, then (3.25) gives the trivial bound $P_{\mathrm{e}} \leq 1$; this shows that similarly to the TSB (but in contrast to the union bound), the tangential bound and its simplified version do not diverge at low SNR values.

3.2.4 The Divsalar bound

In [50], Divsalar derived a simple upper bound on the ML decoding error probability of linear block codes whose transmission takes place over an AWGN channel. The simplicity of the bound stems from the fact that it is given in closed form, and its calculation does not involve numerical integration and parameter optimization. It is therefore widely used for calculating bounds on the thresholds of turbo-like code ensembles under ML decoding, and for assessing the decoding error probability of linear block codes which are communicated over a binary-input AWGN channel (see, e.g., [1, 50, 52, 102, 103, 117, 118, 123, 174, 183]).

In the following, we present the Divsalar bound [50]. In the derivation of the bound here, we prove the final expressions of the Chernoff bounds in [50, Eqs. (14)–(17)] (their proofs are missing in [50]); on the other hand, we skip the technical details related to the derivation of the final closed form expressions for the optimized parameters in this bound, and refer the reader to [50].

Consider an (n,k) binary linear code $\mathcal{C}$ with rate $R = \frac{k}{n}$, and let $\{S_d\}$ designate its distance spectrum. Suppose the code is BPSK modulated and transmitted over an AWGN channel. Let $\underline{c}^i \in \mathcal{C}$ be an arbitrary codeword, and let $\underline{x}^i$ designate the vector where zeros and ones in the codeword $\underline{c}^i$ are mapped to $+1$ and -1, respectively (due to the BPSK modulation). Let $\underline{c}^i$ and $\underline{y}$ be an arbitrary transmitted codeword and the corresponding n-dimensional vector at the output of the channel, respectively. By scaling the vector $\underline{y} = (y_1, y_2, \ldots, y_n)$, its components satisfy for all j the equality

$$y_j = \gamma \, x^i(j) + n_j$$

where $\gamma \triangleq \sqrt{\frac{2RE_{\mathrm{b}}}{N_0}}$, $x^i(j)$ is the j-th component of $\underline{x}^i$, and n_j is a zero mean and unit-variance Gaussian RV which is due to a scaling of the noise sample. Due to the symmetry of the channel and the linearity of the code, the conditional error probability does not depend on the transmitted codeword, so we assume that the all-zero codeword ($\underline{c}^0$) is the transmitted codeword. An error occurs under ML decoding if there exists $i \in \{1, 2, \ldots, 2^{nR} - 1\}$ such that $\sum_{j=1}^{n} y_j \, x^i(j) > \sum_{j=1}^{n} y_j \, x^0(j)$. Now, let $\mathcal{C}_d$ (for $d = 1, 2, \ldots, n$) designate the subset of all the codewords of $\underline{c} \in \mathcal{C}$ whose Hamming weight is d; the size of the subset $\mathcal{C}_d$ is clearly equal to S_d. We designate by E_d the error event where there exists a codeword $\underline{c} \in \mathcal{C}_d$ which is chosen by the ML decoder in preference to $\underline{c}^0$ (the all-zero codeword). Based on the union bound, the decoding error probability is bounded by $P_{\mathrm{e}} \leq \sum_{d>0} \Pr\left\{E_d | \underline{x}^0\right\}$. Divsalar relied on the inequality

$$\Pr\left\{E_d | \underline{x}^0\right\} \leq \Pr\left\{E_d, \underline{y} \in \mathcal{R} | \, \underline{x}^0\right\} + \Pr\left\{\underline{y} \notin \mathcal{R} | \, \underline{x}^0\right\} \tag{3.26}$$

which was used earlier for the derivation of the TSB and the tangential bound, and where (3.26) holds for any region $\mathcal{R}$ in the n-dimensional observation space. Referring to Fig. 3.2, Divsalar defined the region $\mathcal{R}$ to be an n-dimensional sphere with radius $\sqrt{nR^2}$ (R is later optimized). The center of that sphere is at $\eta\gamma\underline{x}^0$, a point along the line connecting

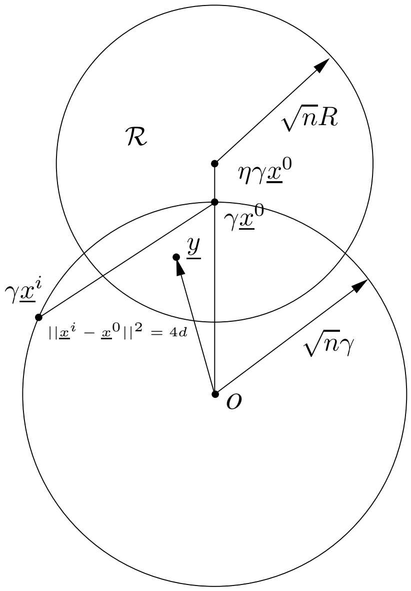

Fig. 3.2 The geometric interpretation of the Divsalar bound [50].

the origin to the transmitted codeword $\underline{x}^0$ (see Fig. 3.2). Based on the union bound and the choice of $\mathcal{R}$ in Fig. 3.2, we obtain the inequality

$$\begin{aligned}\Pr\left\{E_d|\underline{x}^0\right\} \\ &\leq \sum_{\underline{c}^i\in\mathcal{C}_d} \Pr\left\{\sum_{j=1}^n y_j x^i(j) \geq \sum_{j=1}^n y_j x^0(j), \sum_{j=1}^n \big(y_j - \eta\gamma x^0(j)\big)^2 \leq nR^2|\underline{x}^0\right\} \\ &\quad + \Pr\left\{\sum_{j=1}^n \big(y_j - \eta\gamma x^0(j)\big)^2 > nR^2|\underline{x}^0\right\}. \qquad (3.27)\end{aligned}$$

Let

$$Z = \sum_{j=1}^n y_j \left(x^i(j) - x^0(j)\right), \quad W = \sum_{j=1}^n \big(y_j - \eta\gamma x^0(j)\big)^2 - nR^2$$

then, based on the Chernoff bounds

$$\Pr\{Z \geq 0, W \leq 0\} \leq E\left\{e^{tZ+rW}\right\}, \quad \forall\, t \geq 0, \quad r \leq 0 \qquad (3.28)$$

and

$$\Pr\left\{W > 0\right\} \leq E\left\{e^{sW}\right\}, \quad \forall\, s \geq 0 \tag{3.29}$$

Divsalar obtained in [50] Chernoff upper bounds for the two terms in the RHS of (3.27). We derive here the two Chernoff bounds. To this end, one relies on the equality which states that if $X \sim N(m,\sigma^2)$, then

$$E\left[e^{sX^2}\right] = \frac{e^{\frac{m^2 s}{1-2\sigma^2 s}}}{\sqrt{1-2\sigma^2 s}}\,, \quad \forall\, s < \frac{1}{2\sigma^2}. \tag{3.30}$$

Since it is assumed that the all-zero codeword is transmitted, then $\underline{x}^0 = (1,1,\ldots,1)$, and from (3.29) and (3.30) we obtain that

$$\begin{aligned}
&\Pr\left\{\sum_{j=1}^{n}\left(y_j - \eta\gamma x^0(j)\right)^2 > nR^2 \mid \underline{x}^0\right\} \\
&\quad \leq E\left[e^{s\left(\sum_{j=1}^{n}\left(y_j-\eta\gamma x^0(j)\right)^2 - nR^2\right)} \mid \underline{x}^0\right] \qquad s \geq 0 \\
&\quad = e^{-nsR^2}\,\left(f_1(\gamma,s,\eta)\right)^n \triangleq e^{-nsR^2}\,A\,, \quad 0 \leq s < \frac{1}{2}
\end{aligned} \tag{3.31}$$

where

$$f_1(\gamma,s,\eta) = \frac{e^{\frac{(1-\eta)^2\gamma^2 s}{1-2s}}}{\sqrt{1-2s}}\;. \tag{3.32}$$

For $\underline{c}^i \in \mathcal{C}_d$, there are d indices $j \in \{1,2,\ldots,n\}$ where $x^i(j) = -1$, and for the other indices of j: $x^i(j) = 1$. We therefore obtain from the Chernoff bound in (3.28) and the moment generating function in (3.30) that for an arbitrary codeword $\underline{c}^i \in \mathcal{C}_d$, and for arbitrary $t \geq 0$ and $r \leq 0$

$$\begin{aligned}
&\Pr\left\{\sum_{j=1}^{n} y_j\, x^i(j) \geq \sum_{j=1}^{n} y_j\, x^0(j),\ \sum_{j=1}^{n}\left(y_j - \eta\gamma x^0(j)\right)^2 \leq nR^2 \mid \underline{x}^0\right\} \\
&\quad \leq E\left[e^{t\left(\sum_{j=1}^{n} y_j x^i(j) - \sum_{j=1}^{n} y_j x^0(j)\right) + r\left(\sum_{j=1}^{n}\left(y_j-\eta\gamma x^0(j)\right)^2 - nR^2\right)} \mid \underline{x}^0\right] \\
&\quad = e^{-nrR^2}\,\left(f_1(\gamma,r,\eta)\right)^{n-d}\ \left(\frac{e^{\frac{r}{1-2r}\left[(1-\eta)\gamma-\frac{t}{r}\right]^2 - 2\eta\gamma t - \frac{t^2}{r}}}{\sqrt{1-2r}}\right)^d.
\end{aligned} \tag{3.33}$$

The minimization of the exponent in (3.33) with respect to t gives that

$$t = \frac{\gamma}{2}\,(1 - 2r\eta)$$

(t is indeed non-negative, since $r \leq 0$), so the substitution of t in (3.33) yields that for all codewords $\underline{c}^i \in \mathcal{C}_d$

$$\begin{aligned}
&\Pr\left\{\sum_{j=1}^{n} y_j\, x^i(j) \geq \sum_{j=1}^{n} y_j\, x^0(j),\ \sum_{j=1}^{n}\big(y_j - \eta\gamma x^0(j)\big)^2 \leq nR^2 \mid \underline{x}^0\right\}\\
&\quad\leq \Big(f_1(\gamma,r,\eta)\Big)^{n-d}\,\Big(f_2(\gamma,r,\eta)\Big)^{d}\, e^{-nrR^2} \qquad (3.34)
\end{aligned}$$

where the function f_1 is introduced in (3.32) and

$$f_2(\gamma,r,\eta) = \frac{e^{-\frac{\gamma^2(1-2r\eta^2)}{2}}}{\sqrt{1-2r}}\ . \qquad (3.35)$$

Eqs. (3.34) and (3.35) yield therefore the inequality

$$\begin{aligned}
&\sum_{\underline{c}^i\in\mathcal{C}_d}\Pr\left\{\sum_{j=1}^{n} y_j\, x^i(j) \geq \sum_{j=1}^{n} y_j\, x^0(j),\ \sum_{j=1}^{n}\big(y_j - \eta\gamma x^0(j)\big)^2 \leq nR^2 \mid \underline{x}^0\right\}\\
&\leq S_d\Big(f_1(\gamma,r,\eta)\Big)^{n-d}\,\Big(f_2(\gamma,r,\eta)\Big)^{d}\, e^{-nrR^2} \triangleq e^{-nrR^2}B, \quad r \leq 0. \quad (3.36)
\end{aligned}$$

From (3.27), (3.31) and (3.36), the minimization of the resulting upper bound on $\Pr\left\{E_d|\underline{x}^0\right\}$ with respect to R gives

$$e^{-nR^2} = \left(-\frac{Br}{As}\right)^{\frac{1}{s-r}}$$

and the substitution of this optimized value in this upper bound gives

$$\Pr\left\{E_d|\underline{x}^0\right\} \leq 2^{h\left(\frac{s}{s-r}\right)}\, A^{-\frac{r}{s-r}}\, B^{\frac{s}{s-r}}\ , \quad 0 < s < \frac{1}{2},\ r \leq 0$$

where $h(x) = -x\log_2(x) - (1-x)\log_2(1-x)$ is the binary entropy function to the base 2. Instead of the parameters s, r and η, Divsalar introduced in [50] the three new parameters

$$\rho = \frac{s}{s-r},\ \beta = \rho(1-2r),\ \xi = \rho(1-2r\eta) \quad (0 \leq \rho \leq 1,\ \beta \geq 0,\ \xi \geq 0). \qquad (3.37)$$

Let us define $c \triangleq \frac{E_{\mathrm{s}}}{N_0}$, the normalized Hamming weight as $\delta \triangleq \frac{d}{n}$, and the growth rate of the distance spectrum as $r_n(\delta) \triangleq \frac{\ln S_d}{n}$. The resulting upper bound gets the form

$$\Pr\left\{E_d|\underline{x}^0\right\} \leq 2^{h(\rho)}\, e^{-nE(c,\delta,\beta,\rho,\xi)}$$

where

$$\begin{aligned} E(c,\delta,\beta,\rho,\xi) &= -\rho\, r_n(\delta) - \frac{\rho}{2}\ln\left(\frac{\rho}{\beta}\right) - \frac{1-\rho}{2}\ln\left(\frac{1-\rho}{1-\beta}\right) \\ &\quad + c\left[1-(1-\delta)\,\frac{\xi^2}{\beta} - \frac{(1-\xi)^2}{1-\beta}\right]. \end{aligned} \tag{3.38}$$

The parameters ρ, ξ and β were optimized in [50, pp. 6–7] in order to get the maximal error exponent in (3.38) (and hence, the tightest upper bound within this form). Their optimal values are

$$\rho = \frac{\beta}{\beta + (1-\beta)e^{2r_n(\delta)}}\,, \qquad \xi = \frac{\beta}{\beta + (1-\beta)(1-\delta)}$$

where the optimal value of β is given by

$$\begin{aligned} \beta &= \sqrt{\frac{c(1-\delta)}{\delta}\,\frac{2}{1-e^{-2r_n(\delta)}} + \left(\frac{1-\delta}{\delta}\right)^2\left[(1+c)^2-1\right]} \\ &\quad - \frac{1-\delta}{\delta}\cdot(1+c) \end{aligned} \tag{3.39}$$

and $c \triangleq \frac{E_{\mathrm{s}}}{N_0} = \frac{RE_{\mathrm{b}}}{N_0}$.

Let $d_{\min}$ and $d_{\max}$ designate, respectively, the minimal and the maximal Hamming weight (d) so that $S_d \neq 0$. The final form of the Divsalar bound on the block error probability is

$$P_{\mathrm{e}} \leq \sum_{d=d_{\min}}^{d_{\max}} \min\left\{e^{-nE(\delta,\beta,\frac{E_{\mathrm{s}}}{N_0})}, S_d\, Q\left(\sqrt{\frac{2dE_{\mathrm{s}}}{N_0}}\right)\right\} \tag{3.40}$$

where

$$\begin{aligned} E(\delta,\beta,\frac{E_{\mathrm{s}}}{N_0}) &= -r_n(\delta) + \frac{1}{2}\ln\left(\beta + (1-\beta)e^{2r_n(\delta)}\right) \\ &\quad + \frac{\beta\delta}{1-(1-\beta)\delta}\,\frac{E_{\mathrm{s}}}{N_0} \end{aligned} \tag{3.41}$$

and β is given in (3.39). We note that if $\beta = 1$, the bound in (3.40) reduces to the union bound, so since β in (3.39) is the optimized value, then it follows immediately that the Divsalar bound is inherently tighter than the union bound. To compute an upper bound on the bit error probability, the distance spectrum S_d is replaced by $\sum_{w=1}^{nR} \left\{ \left(\frac{w}{nR}\right) A_{w,d} \right\}$ where $A_{w,d}$ designates the number of codewords whose Hamming weight is equal to d, and which are encoded by information bits of Hamming weight w (i.e., $A_{w,d}$ designates the input-output weight enumerator of the code $\mathcal{C}$).

The Divsalar bound provides a closed-form upper bound on the threshold of the signal to noise ratio which yields vanishing block error probability under ML decoding. By letting the block length tend to infinity, the upper bound on the $\frac{E_\text{b}}{N_0}$ threshold [50, Eq. (33)] is given by

$$\left(\frac{E_\text{b}}{N_0}\right)_\text{threshold} \leq \frac{1}{R} \max_{0<\delta<1} \frac{(1-\delta)(1-e^{-2r(\delta)})}{2\delta} \tag{3.42}$$

where $r(\delta) = \lim_{n\to\infty} r_n(\delta)$ designates the asymptotic growth rate of the distance spectrum.

3.2.5 Sphere upper bound

Herzberg and Poltyrev derived the sphere upper bound [90, Section 2A]. This bound relies on the basic inequality in (3.1) where the region $\mathcal{R}$ is chosen to be a sphere whose center coincides with the signal point $(\underline{s}^0)$ which represents the transmitted signal. It is therefore a particular instance of the Divsalar bound (see Section 3.2.4) where the value of η in Fig. 3.2 is set to one (i.e., the case where the center of the sphere coincides with the signal point which represents the transmitted codeword).

Let $\underline{z}$ designate the AWGN vector, then we obtain from the (3.1) and the union bound with respect to the sphere region that

$$\begin{aligned} P_\text{e} &\leq \Pr(\text{error},\ ||\underline{z}|| \leq r) + \Pr(||\underline{z}|| > r) \\ &\leq \sum_k \Big\{ S_k \cdot \Pr(E_k,\ ||\underline{z}|| \leq r) \Big\} + \Pr(||\underline{z}|| > r) \end{aligned} \tag{3.43}$$

where E_k is the error event at the output of the ML decoder for which the Euclidean distance between the signal corresponding to the ML decision and the transmitted signal $\underline{s}^0$ is δ_k, and S_k designates the number of signal points at distance δ_k from $\underline{s}^0$. Since the joint probability $\Pr(E_k,\ ||\underline{z}|| \leq r)$ is equal to zero for $r < \frac{\delta_k}{2}$, then the summation in (3.43) can be restricted to values of k such that $r \geq \frac{\delta_k}{2}$. Let $\delta_1, \delta_2, \ldots, \delta_k$ be the sequence of Euclidean distances from $\underline{s}^0$ which are ordered in an increasing order, and let $N(r)$ designate the maximal positive integer k so that $r > \frac{\delta_k}{2}$. Then, we obtain from (3.43) that

$$P_{\mathrm{e}} \leq \min_{r>0} \left\{ \sum_{k=1}^{N(r)} \left\{ S_k \cdot \Pr(E_k,\ ||\underline{z}|| \leq r) \right\} + \Pr(||\underline{z}|| > r) \right\} \tag{3.44}$$

Referring to Fig. 3.2, the sphere bound is obtained for the particular case where $\eta = 1$, but the calculations of the probabilities in the RHS of (3.44) are done exactly, without the need for Chernoff bounds. It is also easily observed that in the limit where $r \to \infty$, the sphere bound in (3.43) coincides with the union bound, so the sphere bound with the optimized radius r in (3.44) cannot be worse than the union bound. On the other hand, in the limit where $r \to 0$, the sphere bound becomes the trivial bound of unity (where on the other hand, the union bound may exceed unity at low signal to noise ratios).

Let z_1 be the component of the noise vector $\underline{z}$ in the direction of the line connecting $\underline{s}^0$ with $\underline{s}^k$, and let $Y \triangleq \sum_{i=1}^n z_i^2$ be the square of the magnitude of the noise vector $\underline{z}$. Due to the spherical symmetry of the AWGN, we obtain that

$$\begin{aligned} \Pr(E_k,\ ||\underline{z}|| \leq r) \Big\} &= \Pr\left(\frac{\delta_k}{2} \leq z_1 \leq r,\ Y \leq r^2\right) \\ &= \int_0^{r^2} \int_{\frac{\delta_k}{2}}^{r} p_{z_1,Y}(z_1, y) dz_1 dy \end{aligned} \tag{3.45}$$

where

$$p_{z_1,Y}(z_1, y) = \frac{(y - z_1^2)^{\frac{n-3}{2}} e^{-\frac{y}{2\sigma^2}}\, U(y - z_1^2)}{\sqrt{\pi} 2^{\frac{n}{2}} \sigma^n \Gamma\left(\frac{n-1}{2}\right)} \tag{3.46}$$

is the joint probability density function of z_1 and Y; since z_1 and $Y - z_1^2 = \sum_{i=2}^n z_i^2$ are i.i.d. the pdf in (3.46) is calculated by multiplying the normal pdf of z_1 and the pdf of $Y - z_1^2$ (which is a χ^2 distribution with $n-1$ degrees of freedom). The functions U and Γ in

(3.46) are the unit step function and the complete Gamma function (see (3.3)), respectively. Similarly, since the random variable Y is χ^2 distributed with n degrees of freedom, then we obtain that the second probability in the RHS of (3.44) is equal to

$$\begin{aligned}\Pr\left(||\underline{z}|| > r\right) &= \Pr\left(Y > r^2\right)\\ &= \int_{r^2}^{\infty} \frac{y^{\frac{n-2}{2}}\, e^{-\frac{y}{2\sigma^2}}}{2^{\frac{n}{2}}\, \sigma^n\, \Gamma\left(\frac{n}{2}\right)}\, dy\\ &= 1 - \gamma\left(\frac{n}{2}, \frac{r^2}{2\sigma^2}\right) \qquad (3.47)\end{aligned}$$

where γ designates the incomplete Gamma function introduced in (3.7). Finally, the sphere upper bound follows by the substitution of (3.45)–(3.47) into (3.44). The optimization of the radius r in the sphere upper bound (3.44) is obtained by a numerical solution of the equation

$$\sum_{k=1}^{N(r)} S_k \int_0^{\theta_k} \sin^{n-2}\phi\, d\phi = \frac{\sqrt{\pi}\,\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \qquad (3.48)$$

where

$$\theta_k \triangleq \cos^{-1}\left(\frac{\delta_k}{2r}\right), \quad k = 1,2,\ldots,N(r).$$

3.2.6 The Engdahl and Zigangirov bound

In [67], Engdahl and Zigangirov derived new upper bounds on the bit and burst error probabilities for a convolutional code transmitted over an AWGN channel. For the analysis, they consider simple error events (i.e., a trellis which does not contain any shorter trellises which start and end in the all-zero state). The derivation of their bounds is based on partitioning the set of simple error events according to the length l of the trellis path. Considering an antipodal modulation of the binary symbols (where 0 and 1 are mapped to $+\sqrt{E_{\mathrm{s}}}$ and $-\sqrt{E_{\mathrm{s}}}$, respectively), it is assumed w.o.l.g. that the all zero sequence is transmitted. Let us assume that the convolutional code is of rate $R_{\mathrm{c}} = \frac{b}{c}$, and let $\underline{y}_t = (y_t^{(1)}, y_t^{(2)}, \ldots, y_t^{(c)})$ designate the received symbols at the output of the

AWGN channel at time t. In order to analyze the probability of an error event of length l, we introduce the random variable

$$X^{(l)} = \sum_{t=0}^{l-1} \sum_{i=1}^{c} y_t^{(i)}~, \quad l = m+1, m+2, \dots$$

where m is the memory length of the convolutional encoder (note that $m+1$ is the shortest possible error event). Since the all-zero codeword is transmitted, $X^{(l)}$ is Gaussian distributed, and $X^{(l)} \sim N(cl\sqrt{E_\mathrm{s}}, \frac{clN_0}{2})$. Let $\mathcal{D}$ be the set of indices (t,i) referring to the code symbols in the relevant error event which are non-zero (where $t = 0, 1, \dots, l-1$ and $i = 1, 2, \dots, c$), and let

$$Y = \sum_{(t,i)\in\mathcal{D}} y_t^{(i)}.$$

The ML decoder (a Viterbi decoder) favors the error event sequence referring to the above indices $\{t,i\}$, and therefore makes an erroneous decision if and only if $Y < 0$. The Engdahl and Zigangirov bounds on the bit and burst error probabilities of convolutional codes are based on the basic inequality in (3.1) where the "good region" $\mathcal{R}$ which in general depends on the length l corresponds to the case where the random variable $X^{(l)}$ is above a certain threshold u_l. This value u_l is later optimized individually for every l.

Since we consider the Gaussian channel, the conditional and un-conditional pairwise error probabilities for an arbitrary value of l (conditioned on the value of the random variable $X^{(l)}$) are calculated exactly without the need for Chernoff bounds. The Engdahl and Zigangirov bounds on the burst and bit error probabilities are stated in [67, Theorems 1 and 2], and these bounds are uniformly tighter than the respective union bounds on the burst and bit error probabilities.

We note that Craig's identity (2.9) for the Q-function (see [40]) can be very useful in expressing the Engdahl and Zigangirov bound in terms of the weight enumerator of the convolutional code, if such a function is available. This approach provides an exact upper bound, instead of the truncation which is used for the numerical calculation of the bound in [67, Section 4]. In this respect, we note that McEliece derived a general technique for calculating the weight enumerators of convolutional codes (see [127]). Finally, we note that the adaptation of the bounding technique of Engdahl and Zigangirov in order to obtain upper bounds on the block and bit error probabilities of linear block codes is straightforward, and even leads to simpler bounds; for linear block codes, there is no need to sum over all the possible lengths of unmerged paths in the trellis diagram (as all the codewords have the same block length). For linear block codes, it yields upper bounds which are subject to one parameter optimization, and in fact, this optimized parameter can be computed numerically as the single root of a related optimization equation.

Consider the case where the codewords of a binary linear block code are BPSK modulated and transmitted over a binary-input AWGN channel. The concept of the derivation of the Engdahl and Zigangirov bound in [67] yields a conceptually similar bound where the region $\mathcal{R}$ in (3.1) forms a *plane* in the n-dimensional Euclidean space, where the "good region" is associated with the case where the correlation between the received vector and the transmitted codeword exceeds a certain threshold (to be optimized in order to obtain the tightest bound within this form). The resulting final form of this bound is the following:

$$\begin{aligned} P_{\mathrm{e}} \leq \sum_{d=1}^{n} \Bigg\{ S_d \sqrt{\frac{E_{\mathrm{s}}}{\pi n N_0}} \int_{\eta}^{\infty} Q\left(\sqrt{\frac{2E_{\mathrm{s}}}{N_0}\,\frac{d}{n(n-d)}}\;t\right) \\ \cdot \exp\left(-\frac{E_{\mathrm{s}}}{N_0}\,\frac{(t-n)^2}{t}\right) dt \Bigg\} + Q\left(\sqrt{\frac{2E_{\mathrm{s}}}{N_0}}\,\frac{n-\eta}{\sqrt{n}}\right) \end{aligned} \tag{3.49}$$

where the optimal value of the parameter η, associated with the threshold for the correlation between the received vector and the transmitted codeword which determines in this case the region $\mathcal{R}$ in (3.1), is calculated by solving numerically the equation

$$\sum_{d=1}^{n} S_d Q\left(\sqrt{\frac{2E_{\mathrm{s}}}{N_0}\,\frac{d}{n(n-d)}}\;\eta\right) = 1. \tag{3.50}$$

The last equation is equivalent to [67, Eq. (27)]. We note that in [67], Engdahl and Zigangirov obtained (3.50) in a different approach, but their choice for η happens to yield the optimal value.

Since the geometrical regions $\mathcal{R}$ in (3.1) which correspond to the Engdahl and Zigangirov bound and the tangential bound are both planes in the n-dimensional Euclidean space whose free parameter is optimized in order to minimize the resulting upper bound, they actually can be shown to yield an equivalent bound [120]. The novelty in the derivation of the Engdahl and Zigangirov bound in [67] is in the adaptation of their bounding technique for obtaining upper bounds on the bit error probability and burst error probability of ML decoded convolutional codes (where on the other hand, the tangential bound was derived by Berlekamp [22] as an upper bound on the block error probability of binary linear block codes).

3.2.7 The 1966 Viterbi bound

The 1966 Viterbi bound was derived in [206, Section 8]. The general idea of this bound is that the decoding error probability of a code becomes worse by increasing the correlation between the codewords. Hence, if we define $\rho_{\max}$ to be the maximal correlation between any two codewords of the code, then the decoding error probability of the code is upper bounded by the error probability which results in if the correlation between any two codewords of the code was equal to $\rho_{\max}$. If the communication takes place over an AWGN channel, then we get the following bound on the decoding error probability

$$P_{\mathrm{e}} \leq 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} \cdot \left[1 - Q\left(x + \sqrt{\frac{2nRE_{\mathrm{b}}(1-\rho_{\max})}{N_0}}\right)\right]^{M-1} dx \tag{3.51}$$

where n and R are the block length and the rate of the code, and $M = 2^{nR}$ is the number of codewords. If we assume that the codewords of a binary linear block code $\mathcal{C}$ are BPSK modulated, then the maximal correlation between any two codewords of $\mathcal{C}$ is directly linked to the minimum distance ($d_{\min}$) of the code, and satisfies the simple relation

$$\rho_{\max} = 1 - \frac{2d_{\min}}{n}. \tag{3.52}$$

The bound is actually the exact decoding error probability for binary linear block codes with fixed composition (i.e., codes whose all codewords (except the all-zero codeword) have constant Hamming weight) and in particular, it gives the exact decoding error probability for bi-orthogonal codes and the dual of Hamming codes.

We note that if X is a zero mean, unit-variance, normal random variable, then the following equality holds (see [205, Eq. (3.66)])

$$E\Big[Q(\mu + \lambda X)\Big] = Q\left(\frac{\mu}{\sqrt{1+\lambda^2}}\right) \,, \quad \forall\, \mu, \lambda \in \mathbb{R}$$

and by setting $\lambda = 1$ in the latter equality, we obtain that

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}\, Q(x+\mu)\, dx = Q\left(\frac{\mu}{\sqrt{2}}\right) \,, \quad \forall\, \mu \in \mathbb{R}.$$

Since from Jensen's inequality

$$E\Big[1 - (1-y)^{M-1}\Big] \leq 1 - \Big(1 - E[y]\Big)^{M-1}$$

then, similarly to the approach in [99], we obtain from the Viterbi bound (3.51) a looser but a simplified upper bound on the ML decoding error probability:

$$P_{\mathrm{e}} \leq 1 - \left[1 - Q\left(\sqrt{\frac{nRE_{\mathrm{b}}(1-\rho_{\max})}{N_0}}\right)\right]^{M-1} . \qquad (3.53)$$

The bound in (3.53) is useful for the case of a small error probability, and it circumvents the need for the numerical integration in (3.51).

3.2.8 Hughes' bound

In [98], Hughes derived an upper bound on the ML decoding error probability for *any signal set* which is communicated over the AWGN channel. As opposed to upper bounds presented earlier in this section, note that the transmitted signals in this case are not necessarily expected to have equal energy, but are only assumed to be equally probable. Let the signal set be represented by the signal points $\underline{s}^0, \underline{s}^1, \ldots, \underline{s}^{M-1}$ where M designates the cardinality of the signal set. The basic idea in the derivation of the bound stems from the geometry of the decision regions under ML decoding; each decision region is a polyhedron which is formed by at most $M-1$ supporting hyper-planes. Now, the complementary of the decision region which corresponds to the transmitted signal is decomposed into a disjoint union of $M-1$ truncated polyhedral cones with vertices at the transmitted signal point ($\underline{s}^0$) (see [98, Section 2]). Then, it is shown that the conditional error probability is increased by replacing these $M-1$ polyhedral cones by $M-1$ circular cones with the same solid angles, and with positive axes which are coincident with the rays from $\underline{s}^0$ through $\underline{s}^i$ (where $i = 1,2,\ldots,M-1$). Since the exact geometry of the decision regions is not available in general, then the upper bound on the conditional ML decoding error probability is finally obtained by maximizing over all feasible solid angles of the circular cones.

In order to state the bound explicitly, let n be the number of coordinates of each signal point $\underline{s}^i$ $(i = 0,1,\ldots,M-1)$, and let us define the function

$$B_n(\theta,x) \triangleq \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{n-1}{2}\right)} \int_0^{\theta} \sin^{n-2}\phi \cdot \left(1 - \gamma\left(\frac{n}{2}, \frac{x^2 \sec^2\phi}{2}\right)\right) d\phi \tag{3.54}$$

where $\Gamma(x)$ designates the Gamma function in (3.3), and $\gamma(a,x)$ designates the incomplete Gamma function in (3.7). Intuitively, $B_n(\theta,x)$ represents the probability that an n-tuple of i.i.d. Gaussian random variables with zero mean and unit variance, lies in a circular cone of half-angle θ truncated at distance x from the origin (we refer here to the resulting infinite circular cone whose axis is the ray connecting the two signal points $\underline{s}^0$ and $\underline{s}^i$, and which starts at distance x from $\underline{s}^0$). Useful recursive formulas for the calculation of the function $B_n(\theta,x)$ are provided in [98, Section 3].

Let $\delta_i \triangleq ||\underline{s}^i - \underline{s}^0||$ be the Euclidean distance between $\underline{s}^i$ and $\underline{s}^0$ (where $i = 1,2,\ldots,M-1$). The final form of the upper bound on the conditional error probability is stated in [98, Proposition 2], and it reads:

$$P_{\mathrm{e}|0} \leq \sum_{i=1}^{M-1} A_n\left(\frac{\delta_i}{\alpha_0}, \frac{\delta_i}{2\sigma}\right) \tag{3.55}$$

where $\alpha = \alpha_0$ is the unique positive solution of the equation

$$\sum_{i=1}^{M-1} A_n\left(\frac{\delta_i}{\alpha_0}, 0\right) = 1 \tag{3.56}$$

and

$$A_n(x,y) \triangleq B_n\Big(\Theta(x), y\Big), \qquad \Theta(x) \triangleq \begin{cases} \cos^{-1}(x) \ , \ |x| \le 1 \ , \\ \\ 0 \qquad\qquad , \ |x| > 1 \ . \end{cases}$$

A nice property which follows easily from (3.56) is that the value of α_0 does not depend on the noise variance σ^2, so the value of α_0 needs to be calculated only one time for a particular code. It is also noted in [98] that any value of α for which $\sum_{i=1}^{M-1} A_n\left(\frac{\delta_i}{\alpha}, 0\right) \ge 1$ yields a valid upper bound in (3.55).

The bound of Hughes is always tighter than the union bound. Hughes noted in [98] that his bound provides a significant improvement over the union bound if $M \gg n$ (i.e., when the cardinality of the signal set is much larger than the block length of the code). In light of this, the improvement of his bound over the union bound seems to be primarily useful for lattice codes which are densely packed (the reader is referred to the examples in [98, Section 4]).

3.2.9 Voronoi regions for binary linear block codes and expurgated bounds

Consider a set of M points in $\mathbb{R}^n$, $\mathcal{C} = \{\underline{c}^1, \underline{c}^2, \ldots, \underline{c}^M\}$, and the distance measure $d(\underline{u}, \underline{v}) = ||\underline{u} - \underline{v}||$. If all vectors in $\mathbb{R}^n$ are grouped together according to which of the points in $\mathcal{C}$ they are closest to, then the Euclidean space $\mathbb{R}^n$ is partitioned into Voronoi regions. The Voronoi region of a point $\underline{c}^i \in \mathcal{C}$ (say Ω_i) is defined as the set of vectors in $\mathbb{R}^n$ which are closest to this point, i.e.,

$$\Omega_i \triangleq \left\{ \underline{x} \in \mathbb{R}^n \mid d(\underline{x}, \underline{c}^i) \le d(\underline{x}, \underline{c}), \ \forall \ \underline{c} \in \mathcal{C} \right\} \ . \tag{3.57}$$

In case of a tie between two or more points in $\mathcal{C}$, the vectors belong to more than one Voronoi region, but no vector belongs to the interior of more than one Voronoi region. The Voronoi regions of a block code govern many aspects of the code performance on an AWGN channel, and they play a central role in the performance analysis of soft-decision decoding algorithms. A Voronoi region, as defined by the set of inequalities in (3.57), is the intersection of $M-1$ half-spaces, and is therefore a convex region in n dimensions. It has usually fewer facets than

M, which means that part of the inequalities in (3.57) are redundant. In this case, the region Ω_i can be represented by a subset of these inequalities, i.e., there exists a subset $\mathcal{N}_i \subseteq \mathcal{C}$ such that

$$\Omega_i \triangleq \left\{ \underline{x} \in \mathbb{R}^n \mid d(\underline{x}, \underline{c}^i) \le d(\underline{x}, \underline{c}), \, \forall \, \underline{c} \in \mathcal{N}_i \right\} \tag{3.58}$$

and the *minimal set* $\mathcal{N}_i$ for which the RHS of (3.58) is equal to the Voronoi region in (3.57) is called *the set of the Voronoi neighbors* of $\underline{c}^i$. An equivalent definition of Voronoi neighbors which is given in [2] states that the two points $\underline{c}^i, \underline{c}^j \in \mathcal{C}$ are neighbors if

$$\exists \, \underline{x} \in \mathbb{R}^n : d(\underline{x}, \underline{c}^i) = d(\underline{x}, \underline{c}^j) < d(\underline{x}, \underline{c}) \, , \, \forall \, \underline{c} \in \mathcal{C} \backslash \{\underline{c}^i, \underline{c}^j\} \tag{3.59}$$

which means that the Voronoi regions of $\underline{c}^i$ and $\underline{c}^j$ have an $(n-1)$-dimensional facet in common. A related concept which is considered in [2] is *Gabriel neighborhood* which depends in general on a looser condition. These two points are said to be Gabriel neighbors if the condition in (3.59) is satisfied for the vector $\underline{x}$ which is the halfway between $\underline{c}^i$ and $\underline{c}^j$, i.e., if

$$d(\underline{m}^{i,j}, \underline{c}^i) = d(\underline{m}^{i,j}, \underline{c}^j) < d(\underline{m}^{i,j}, \underline{c}) \, , \, \forall \, \underline{c} \in \mathcal{C} \backslash \{\underline{c}^i, \underline{c}^j\} \tag{3.60}$$

where $\underline{m}^{i,j} \triangleq \frac{1}{2} \left(\underline{c}^i + \underline{c}^j \right)$. The geometrical interpretation of a pair of Gabriel neighbors is that the straight line connecting these two points goes directly from one Voronoi region to the other, without touching a third Voronoi region. By comparing (3.59) and (3.60), it follows that Gabriel neighbors are also Voronoi neighbors. The central theorem in [2] states that all Voronoi neighbors of a codeword in a binary linear block code are also its Gabriel neighbors. The proof of this theorem relies on both the linearity of the code and its binary alphabet. Next, it is shown in [2] that for a binary code, the condition for two codewords to be Gabriel neighbors depends only on the positions where both codewords have the same values; two codewords are Gabriel neighbors if and only if there is no other codeword which has the same values in these positions. Therefore, a codeword $\underline{c}$ is a Gabriel neighbor of the all-zero codeword if and only if there is no other codeword having zeros where $\underline{c}$ has zeros. Since for a binary linear block code whose transmission takes place over an AWGN channel, one can assume without any loss of generality that the all-zero codeword is the transmitted codeword, then we can use the latter property to determine all the Gabriel neighbors of the all-zero codeword. The obvious way to perform this test for the Gabriel neighborhood of an arbitrary codeword $\underline{c}$ to the all-zero codeword is to generate all the codewords in the code, comparing each one of them to $\underline{c}$ and checking if there exists a codeword with zeros in all the positions where $\underline{c}$ has zeros. The computational complexity of this approach is $O(2^k n)$ which is not so feasible for practical codes. A more efficient algorithm for binary linear block codes is presented in the appendix of [2]. The general concept of this algorithm is that instead of generating all codewords, the algorithm finds the number of codewords having zeros in certain positions; it is done by linear row operations of the generator matrix G of the code. This matrix is brought into another valid generator matrix of the same code, and it includes two parts: the last r rows contain codewords with the specified zero pattern, and the first $k - r$ rows of the new generator matrix are such that no linear combination of these rows gives this zero pattern. Then, a total of 2^r codewords, including the all-zero codeword and the codeword $\underline{c}$, contain the specified zero pattern. According to the test above for Gabriel neighborhood, the codeword $\underline{c}$ is a neighbor of the all-zero codeword if and only if $r = 1$. The computational complexity of this algorithm for deciding if a single codeword is a neighbor of the all-zero codeword is $O(n^2 k)$, which forms a considerable reduction in complexity unless the dimension (k) of the code is very small. Therefore, because of the equivalence between Gabriel neighbors and Voronoi neighbors for a binary linear block code, the computational complexity which is involved with this algorithm for the determination of the Voronoi region of the all-zero codeword is $O(2^k \cdot n^2 k)$. A further simplification in the characterization of the Voronoi neighbors of the all-zero codeword stems from the observation in [2] that all the codewords of Hamming weight up to $2d_{\min} - 1$ (where $d_{\min}$ designates the minimum distance of the binary linear block code) are necessarily neighbors of the all-zero codeword. It is also proved in [3] that for a general binary linear block code, all codewords of Hamming weight above $n - k + 1$ are necessarily not neighbors of the all-zero codeword; this yields that for the AWGN channel, the union bound (or any improved upper bound on the

ML decoding error probability which relies on the distance spectrum of the code) can be truncated so that codewords of Hamming weight above $n-k+1$ are not taken into consideration.

The distance spectrum (or weight distribution) of a binary linear code served earlier in this section for the derivation of upper bounds on the ML decoding error probability. Since the set of neighbors of a codeword may be smaller than the total number of codewords, then it is useful to look at the set of neighbors $\mathcal{N}_0$ of the all-zero codeword (see Eq. (3.58)), and define the *local distance spectrum* to be the cardinality of the codewords in the set $\mathcal{N}_0$ as a function of their Hamming weights. For a binary linear block code $\mathcal{C}$, let

$$S_d \triangleq |\{\underline{c} \in \mathcal{C}: \; w_{\mathrm{H}}(\underline{c}) = d\}|, \quad d = 0, \dots, n \tag{3.61}$$

and

$$L_d \triangleq |\{\underline{c} \in \mathcal{N}_0: \; w_{\mathrm{H}}(\underline{c}) = d\}|, \quad d = 0, \dots, n \tag{3.62}$$

designate the *(full) distance spectrum* and the *local distance spectrum* of the code $\mathcal{C}$, respectively, where $w_H(\underline{x})$ stands for the Hamming weight of a vector $\underline{x}$. Clearly, $L_d \leq S_d$ for $d = 0, 1, \dots, n$. Since all codewords of Hamming weight up to $2d_{\min} - 1$ are neighbors of the all-zero codeword, then the equality $L_d = S_d$ holds for $d = 0, 1, 2, \dots, 2d_{\min} - 1$. Finally, since all the codewords of Hamming weight above $n-k+1$ are not neighbors of the all-zero codeword, then $L_d \equiv 0$ for $d = n-k+2, n-k+3, \dots, n$.

Relations between the local distance spectrum of a binary linear block code, its extended code and its even weight subcode are presented in [220]. Using these relations, the local distance spectra of some BCH and Reed-Muller codes have been precisely derived in [220].

The *expurgated union bound* of a binary linear block code which is BPSK modulated and transmitted over the AWGN channel gets the form (see [3, Eq. (10)])

$$P_{\mathrm{e}} \leq \sum_{d=d_{\min}}^{n-k+1} L_d \, Q\left(\sqrt{\frac{2dRE_{\mathrm{b}}}{N_0}}\right). \tag{3.63}$$

Even without knowing the exact set of neighbors of the all-zero codeword, since the expurgated union bound does not include codewords of

Hamming weight above $n - k + 1$, then the error probability is upper-bounded by

$$P_{\mathrm{e}} \leq \sum_{d=d_{\min}}^{n-k+1} S_d \, Q\left(\sqrt{\frac{2dRE_{\mathrm{b}}}{N_0}}\right) \tag{3.64}$$

which is still a better upper bound than the original (i.e., the non-expurgated) union bound.

Considering the Voronoi regions for binary linear block codes, it is natural to ask what is the fraction of codewords of a binary linear block code which are neighbors of a certain codeword (e.g., the all-zero codeword). This question is especially interesting for sufficiently long codes where it is of interest to characterize the asymptotic fraction of the codewords which are neighbors of the all-zero codeword. For an arbitrary binary linear block code $\mathcal{C}$, the fraction of codewords which are neighbors of the all-zero codeword is called the *neighbor ratio*, and following the notation in [3], it is designated here by $\Gamma(\mathcal{C})$. According to [3, Theorem 5], for any binary linear block code $\mathcal{C}$ whose rate satisfies the inequality $R > \frac{1}{2} + \frac{1}{n}$ (i.e., its rate is only slightly above one-half), the neighbor ratio is upper bounded by

$$\Gamma(\mathcal{C}) \leq \left(R - \frac{1}{2} - \frac{1}{n}\right)^{-2} \frac{1}{4n}$$

so if $R > \frac{1}{2}$ and we let the block length tend to infinity, then the neighbor ratio tends to zero. On the other hand, it was proved in [3, Theorem 8] that if $R < \frac{1}{2}$ and $\varepsilon > 0$ (ε can be chosen arbitrarily small), then the fraction of binary linear block codes of length n and rate R whose neighbor ratio is above $1 - \varepsilon$ tends to 1 as $n \to \infty$. These observations show the nice property that $R = \frac{1}{2}$ forms a threshold with respect to the asymptotic neighbor ratio of binary linear block codes; for code rates above this threshold, the neighbor ratio tends asymptotically to zero, and for code rates below this threshold, the neighbor ratio tends asymptotically to unity with probability 1. Therefore, the expurgated upper bounds on the ML decoding error probability are expected to be especially effective for *high* code rates.

The characterization of the local distance spectra of random linear block codes, Hamming codes and the second-order Reed Muller codes is presented in [6, Theorems 2.2, 2.7 and 2.9]. For the ensemble of random q-ary linear block codes whose parity-check matrices have independent and equiprobable entries, the average local distance spectrum is given in [6, Theorem 2.2], and it reads

$$L_d = \begin{cases} \binom{n}{d} \frac{(q-1)^d}{q^{n-k}} \prod_{i=0}^{d-2} (1 - q^{-(n-k-i)}) & \text{for } d \leq n-k+1, \\ 0 & \text{otherwise.} \end{cases} \tag{3.65}$$

The average distance spectrum of this ensemble can be easily derived. Since the probability that a given vector satisfies a random parity-check equation is $\frac{1}{q}$, then the probability that this vector is contained in a random linear block code with $n-k$ parity-check equations is $q^{-(n-k)}$, and the average distance spectrum of such an ensemble is therefore given by

$$S_d = \binom{n}{d} \frac{(q-1)^d}{q^{n-k}}. \tag{3.66}$$

In the continuation of this section, we consider the effect of such an expurgation on the tightness of the upper bounds on the error probability, and compare the tightness of the expurgated union bound to the expurgated TSB. From these comparisons between the union bound and the TSB with and without expurgation, one can study how much of the gain is due to the expurgation of non-neighbor codewords, and how much is gained due to the improvement in the tangential-sphere bounding technique as compared to the union bound. The effect of both improvements is expected to be more pronounced in the low SNR region, as shown in Figs. 3.3 and 3.4.

In Fig. 3.3, we refer to ensembles of random binary linear block codes with fixed block length (n) and rate (R), and whose parity-check matrices have independent and equiprobable entries. The average *local* distance spectrum of random binary linear block codes is given in (3.65) for $q = 2$, so we rely on this result for the evaluation of the expurgated union bound and the expurgated TSB. From the two plots in Fig. 3.3, it follows that for both bounds (i.e., the union bound and the TSB), the expurgation of non-neighbor codewords of the all-zero codeword provides a certain improvement on their tightness. Consider the ensemble of binary linear block codes of length $n = 100$ and code rate $R = 0.95$

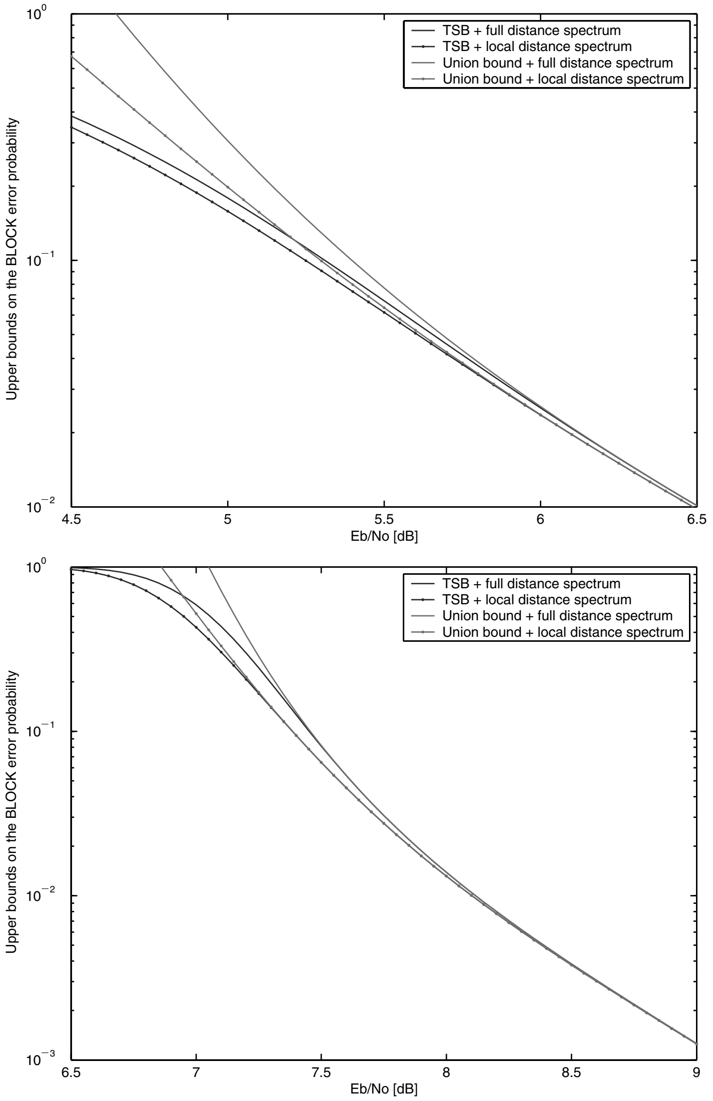

Fig. 3.3 Comparison between upper bounds on the block error probability under soft-decision ML decoding. The performance is evaluated for the ensemble of random binary linear block codes whose parity-check matrices have independent and equiprobable entries. The codes are BPSK modulated, transmitted over a binary-input AWGN channel, and coherently detected. The union bounds are compared to the tangential-sphere bounds (TSBs) with and without expurgation (i.e., with respect to the local and full distance spectrum, respectively). We refer to the ensembles of codes whose block length is $n = 100$ bits and the rate is $R = 0.95$ bits per channel use (upper plot), and the ensemble whose block length is $n = 1000$ bits and the rate is $R = 0.995$ bits per channel use (lower plot).

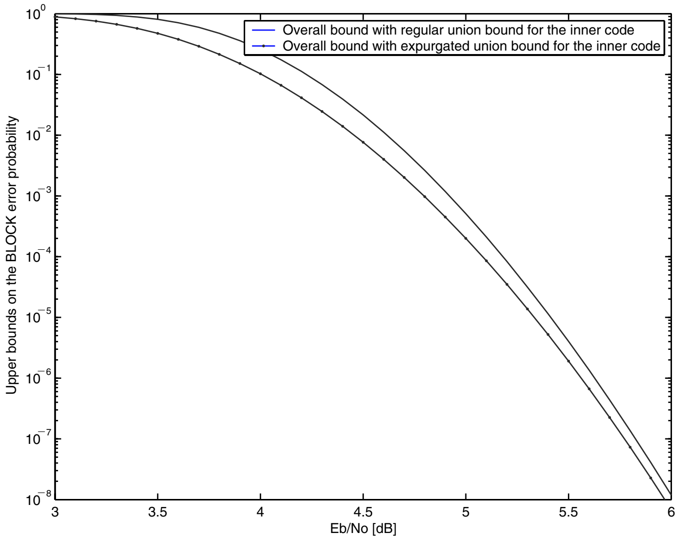

Fig. 3.4 A comparison between two upper bounds which refer to the ensemble performance of serially concatenated codes where the outer code is the primitive (127, 99, 29) RS code, and the inner code is a random binary linear (8, 7) block code. The encoded bits are BPSK modulated, transmitted over the AWGN channel and coherently detected. The decoding is done in two stages: the (8, 7) binary linear block code is soft-decision ML decoded, and the output of its decoder is provided to the RS decoder for a hard-decision ML decoding. The looser bound is based on a union bound with respect to the inner code, and the improvement is achieved by expurgating the union bound.

bits per channel use. For a block error probability of 0.1, the expurgated TSB and the expurgated union bound provide a gain of 0.07 dB and 0.10 dB over their non-expurgated versions, respectively (see the upper plot in Fig. 3.3). At a block error probability of 10^{-2}, the expurgated bounds give a gain of 0.02 dB as compared to the non-expurgated bounds (it is shown in this figure that the TSB and the union bound coincide at a block error probability of 10^{-2}, and the only improvement on the tightness of these bounds is made by the expurgation of the distance spectrum). For the ensemble of random binary linear block codes of length $n = 1000$ and code rate $R = 0.995$ bits per channel use, then even at a high block error probability (e.g., a block error probability of 0.1), the union bound and the TSB coincide (see the lower plot in Fig. 3.3); however, the expurgated bounds provide a gain of 0.08 dB over the non-expurgated bounds at a block error probability of 0.1.

We note that for linear block codes, the expurgation of the distance spectrum yields an improvement on the tightness of the upper bounds only for high rate codes; to this end, for longer binary linear block codes, their rate should be made closer to 1 (in Fig. 3.3, we consider the case where the number of the redundant bits in both ensembles is equal to 5).

Fig. 3.4 presents a possible application of the concept of expurgation. Let us consider an ensemble of serially concatenated codes where the outer code is a primitive Reed-Solomon (RS) code with symbols from the Galois field $GF(2^m)$ (so, its block length is $N = 2^m - 1$ symbols), and the inner code is chosen at random from the ensemble of binary linear block codes which extend every symbol of m bits to a codeword of $m + 1$ bits (we set the dimension of the inner code to m bits, and the block length of this inner code is equal to $m + 1$). Let us choose $m = 7$, so the block length of the primitive RS code is $N = 127$ symbols, and we set the dimension of this code to $K = 99$. In this way, we obtain a 14-error-correcting RS code, and the inner code of the ensemble of serially concatenated codes is chosen at random from the ensemble of (8, 7) binary linear block codes. We assume that the encoded bits are BPSK modulated, transmitted over the AWGN channel and coherently detected. The decoding is done in two stages: the (8, 7) binary linear block code which is soft-decision ML decoded, and then the output of the inner decoder is forwarded to the RS decoder for a hard-decision ML decoding (hence, the RS decoder can correct up to $t = \lfloor \frac{d_{\min}-1}{2} \rfloor = \lfloor \frac{N-K}{2} \rfloor = 14$ symbols). We use here the union bound on the decoding error probability of the inner code which then serves as an upper bound on the a-priori error probability of the symbols of the RS code. The decoding error probability of the considered ensemble of serially concatenated codes therefore satisfies

$$P_{\mathrm{e}} \leq 1 - \sum_{i=0}^{t} \binom{N}{i} p^i (1-p)^{N-i} \tag{3.67}$$

where p designates the value of the union bound which refers to the inner code under soft-decision ML decoding. We compare here the bound in (3.67) with the regular union bound with respect to the inner code, and an improved bound which is obtained by combining (3.67) with the expurgation of the union bound. The local distance spectrum of the ensemble of random binary linear block codes is provided by in (3.65) for $q = 2$, and we use this result (as before) for the calculation of the expurgated union bound. The improved bound in Fig. 3.4, which relies on the expurgated union bound, provides a gain of 0.17 dB, 0.12 dB and 0.10 dB for a block error probability of 10^{-2}, 10^{-3} and 10^{-4}, respectively.

3.2.10 Error exponents of the Chernoff versions of various upper bounds

In his paper [50], Divsalar derived simplified Chernoff versions of some upper bounds on the ML decoding error probability. In Section 4, these bounds are demonstrated to be some particular cases of the 1961 Fano-Gallager bounding technique. These simplified versions include the Chernoff version of the tangential-sphere bound (TSB) of Poltyrev, the Chernoff version of the tangential bound and the Chernoff version of the sphere bound. It is demonstrated in this sub-section that the Chernoff versions of the tangential bound and the sphere bound incur error exponents which are looser than the error exponent of the simple bound of Divsalar, and that the error exponent of the TSB actually coincides with the error exponent of the simple bound of Divsalar [50]. The former result actually indicates that for ensembles of codes with large block lengths, the bound of Divsalar is significantly tighter that the tangential bound and the sphere bound, and the latter result indicates that the TSB and the Divsalar bound possess the same exponential behavior when we let the block length tend to infinity.

3.2.10.1 Error exponent of a simplified TSB and its connection to Divsalar's bound

In [50, Section 4.D], Divsalar derived a simplified TSB which is based on the Chernoff bounding technique, and he showed that the error exponent of the simplified TSB coincides with the error exponent of his bound (see Section 3.2.4 here). The derivation of the simplified TSB and the comparison between its error exponent and the error exponent of the Divsalar bound are outlined in [50, Section 4.D]; due to the importance of this result, we prove it here in detail.

For the derivation of a simplified TSB, we calculate Chernoff bounds on the two types of probabilities in the RHS of (3.6). We note that these probabilities are calculated exactly in Section 3.2.1, but the final form of the TSB in (3.16) involves numerical integrations and also a numerical solution of an associated optimization equation (3.12), so it is therefore not expressed in closed form.

We start now the derivation of a Chernoff version of the TSB. Let us substitute $c \triangleq \frac{E_\mathrm{s}}{N_0}$ and $\eta \triangleq \tan^2(\theta)$. Based on the geometrical interpretation of the TSB in Fig. 3.1 (see Section 3.2.1), we obtain that

$$\begin{aligned}
\Pr\Big(\underline{y} \notin C_n(\theta)\Big) &= \Pr\Big(\textstyle\sum_{i=2}^n z_i^2 > r_{z_1}^2\Big) \\
&\leq E\Big[e^{s'\left(\sum_{i=2}^n z_i^2 - r_{z_1}^2\right)}\Big] \qquad s' \geq 0 \\
&= \textstyle\prod_{i=2}^n E\Big[e^{s' z_i^2}\Big] \cdot E\Big[e^{-s' r_{z_1}^2}\Big] \\
&\overset{\text{(a)}}{=} \left(\tfrac{1}{\sqrt{1-2\sigma^2 s'}}\right)^{n-1} \cdot E\Big[e^{-s' r_{z_1}^2}\Big] \\
&\overset{\text{(b)}}{=} \left(\tfrac{1}{\sqrt{1-2\sigma^2 s'}}\right)^{n-1} \cdot E\Big[e^{-s'\eta\left(z_1 - \sqrt{nE_\mathrm{s}}\right)^2}\Big] \\
&\overset{\text{(c)}}{=} \left(\tfrac{1}{\sqrt{1-2\sigma^2 s'}}\right)^{n-1} \cdot \frac{e^{-\frac{n\eta E_\mathrm{s} s'}{1+2\eta\sigma^2 s'}}}{\sqrt{1+2\eta\sigma^2 s'}} \\
&= \left(\tfrac{1}{\sqrt{1-2s}}\right)^{n-1} \cdot \frac{e^{-\frac{n\eta E_\mathrm{s} s/\sigma^2}{1+2\eta s}}}{\sqrt{1+2\eta s}} \qquad s \triangleq \sigma^2 s' \geq 0 \\
&= \left(\tfrac{1}{\sqrt{1-2s}}\right)^{n-1} \cdot \frac{e^{-\frac{2n\eta cs}{1+2\eta s}}}{\sqrt{1+2\eta s}} \\
&= \sqrt{\tfrac{1-2s}{1+2\eta s}} \left(\tfrac{1}{\sqrt{1-2s}}\right)^n e^{-\frac{2n\eta cs}{1+2\eta s}}
\end{aligned}$$

where equality (a) follows from (3.30) and since $z_i \sim N(0,\sigma^2)$ for $i = 1,2,\ldots,n$ (where $\sigma^2 = \frac{N_0}{2}$), equality (b) follows from (3.2) and the substitution $\eta = \tan^2(\theta)$, and equality (c) follows from (3.30) and since $z_1 - \sqrt{nE_\mathrm{s}} \sim N(-\sqrt{nE_\mathrm{s}}, \frac{\sigma^2}{2})$. We therefore obtain the following Chernoff upper bound on the probability that the received vector $\underline{y}$

falls outside the cone $C_n(\theta)$ in Fig. 3.1

$$\Pr\Big(\underline{y} \notin C_n(\theta)\Big) \leq \sqrt{\frac{1-2s}{1+2\eta s}} \cdot e^{-nE_1(c,s,\eta)} \tag{3.68}$$

where

$$E_1(c,s,\eta) \triangleq \frac{2\eta c s}{1+2\eta s} + \frac{1}{2}\ln(1-2s)\ , \quad 0 \leq s < \frac{1}{2}\ . \tag{3.69}$$

Based on the notation in Section 3.2.1 and Fig. 3.1, the Chernoff bounding technique gives the following upper bound on the joint probability:

$$\begin{aligned}
&\Pr\Big(E_k(z_1), \underline{y} \in C_n(\theta)\Big)\\
&= \Pr\left(\beta_k(z_1) \leq z_2,\ \sum_{i=2}^{n} z_i^2 \leq r_{z_1}^2\right)\\
&= E_{z_1}\left[\Pr\left(\beta_k(z_1) \leq z_2,\ \sum_{i=2}^{n} z_i^2 \leq r_{z_1}^2\ |z_1\right)\right]\\
&\leq E_{z_1}\left[e^{s'(z_2-\beta_k(z_1))+t'(\sum_{i=2}^{n} z_i^2 - r_{z_1}^2)}|z_1\right] \quad s' \geq 0,\ t' \leq 0\\
&= E\Big[e^{-s'\beta_k(z_1)-t' r_{z_1}^2}\Big]\ E\Big[e^{s' z_2 + t' z_2^2}\Big] \prod_{i=3}^{n} E\Big[e^{t' z_i^2}\Big].
\end{aligned} \tag{3.70}$$

Since $z_i \sim N(0,\sigma^2)$, then we obtain from (3.30) that

$$\begin{aligned}
E\Big[e^{s' z_2 + t' z_2^2}\Big] &= e^{-\frac{(s')^2}{4t'}}\ E\Big[e^{t'(z_2+\frac{s'}{2t'})^2}\Big]\\
&= e^{-\frac{(s')^2}{4t'}} \cdot \frac{e^{\frac{\left(\frac{s'}{2t'}\right)^2 t'}{1-2\sigma^2 t'}}}{\sqrt{1-2\sigma^2 t'}}\\
&= \frac{1}{\sqrt{1-2\sigma^2 t'}} \cdot e^{\frac{\sigma^2 (s')^2}{2(1-2\sigma^2 t')}}
\end{aligned} \tag{3.71}$$

and

$$E[e^{t' z_i^2}] = \frac{1}{\sqrt{1-2\sigma^2 t'}} \quad i = 3, 4, \ldots, n,\ \ t' \leq 0. \tag{3.72}$$

By substituting (3.71) and (3.72) into (3.70), we obtain that for $t' \leq 0$ and $s' \geq 0$

$$\Pr\left(E_k(z_1), \underline{y} \in C_n(\theta)\right) \leq E\left[e^{-s'\beta_k(z_1) - t' r_{z_1}^2}\right] \cdot \left(\frac{1}{\sqrt{1-2\sigma^2 t'}}\right)^{n-1} e^{\frac{\sigma^2 (s')^2}{2(1-2\sigma^2 t')}} \qquad (3.73)$$

so it only remains to calculate the statistical expectation with respect to $z_1 \sim N(0, \sigma^2)$ for the term appearing in the RHS of (3.73). Let us define $\delta \triangleq \frac{k}{n}$ as the normalized Hamming weight of a competitive codeword (where we assume without any loss of generality that the all-zero codeword is transmitted). For BPSK modulated signals, the Euclidean distance between the considered pair of codewords is $\delta_k = 2\sqrt{kE_{\mathrm{s}}}$, and with the substitution $\eta \triangleq \tan^2(\theta)$, we obtain from (3.2) that

$$\beta_k(z_1) = \sqrt{\frac{\delta}{1-\delta}}\left(\sqrt{nE_{\mathrm{s}}} - z_1\right), \quad r_{z_1} = \sqrt{\eta}\left(\sqrt{nE_{\mathrm{s}}} - z_1\right). \qquad (3.74)$$

Substituting (3.74) in the RHS of (3.73) gives

$$-s'\beta_k(z_1) - t' r_{z_1}^2 = -t'\eta(z_1 - A)^2 + B$$

where

$$A \triangleq \frac{s'}{2t'\eta}\sqrt{\frac{\delta}{1-\delta}} + \sqrt{nE_{\mathrm{s}}}, \quad B \triangleq \frac{(s')^2}{4t'\eta}\frac{\delta}{1-\delta} \qquad (3.75)$$

and then, the use of (3.30) gives the equality

$$E\left[e^{-s'\beta_k(z_1) - t' r_{z_1}^2}\right] = \frac{e^{B - \frac{A^2 \eta t'}{1+2\eta\sigma^2 t'}}}{\sqrt{1+2\eta\sigma^2 t'}} \quad 1 + 2\eta\sigma^2 t' > 0. \qquad (3.76)$$

From (3.73) and (3.76), we obtain the Chernoff upper bound

$$\Pr\left(E_k(z_1), \underline{y} \in C_n(\theta)\right) \leq \frac{1}{\sqrt{1+2\eta\sigma^2 t'}}\left(\frac{1}{\sqrt{1-2\sigma^2 t'}}\right)^{n-1} \cdot e^{B - \frac{A^2 t' \eta}{1+2\eta\sigma^2 t'} + \frac{\sigma^2 (s')^2}{2(1-2\sigma^2 t')}} \qquad (3.77)$$

where $t' \leq 0$, $s' \geq 0$, $1 + 2\eta\sigma^2 t' > 0$, and the parameters A and B are given in (3.75). We now calculate the optimal value of $s' \geq 0$ which achieves a minimal value for the upper bound in (3.77). By substituting A and B in (3.75) into the exponent in the RHS of (3.77), and introducing the new parameter $v \triangleq \sigma^2 t'$, we obtain that the following equality holds

$$B - \frac{A^2 t' \eta}{1 + 2\eta\sigma^2 t'} + \frac{\sigma^2 (s')^2}{2(1 - 2\sigma^2 t')} = \alpha (s')^2 - \beta s' - \gamma \tag{3.78}$$

where

$$\begin{cases} \alpha \triangleq \frac{\sigma^2}{2} \left(\frac{1}{1+2v\eta} \, \frac{\delta}{1-\delta} + \frac{1}{1-2v} \right) \\ \beta \triangleq \frac{1}{1+2v\eta} \sqrt{\frac{\delta}{1-\delta}} \, \sqrt{nE_{\mathrm{s}}} \\ \gamma \triangleq \frac{nE_{\mathrm{s}}}{\sigma^2} \, \frac{v\eta}{1+2v\eta}. \end{cases} \tag{3.79}$$

We note that since $v \leq 0$ and $1 + 2v\eta > 0$, then $\alpha, \beta > 0$. Hence, the minimal value of the parabola in the RHS of (3.78) is achieved at $s' = \frac{\beta}{2\alpha} \geq 0$, and the value of the parabola at this point is equal to

$$-\frac{\beta^2}{4\alpha} - \gamma = -nc \left(\frac{2v\eta + \frac{\delta}{1-\delta} \, (1-2v)}{1 + 2v\eta + \frac{\delta}{1-\delta} \, (1-2v)} \right), \quad c \triangleq \frac{E_{\mathrm{s}}}{N_0} \, . \tag{3.80}$$

From Eqs. (3.77), (3.78) and (3.80), we obtain the Chernoff upper bound

$$\Pr\Big(E_k(z_1), \underline{y} \in C_n(\theta)\Big) \leq \sqrt{\frac{1-2v}{1+2v\eta}} \, \left(\frac{1}{\sqrt{1-2v}} \right)^n \cdot e^{-nc \cdot \left(\frac{2v\eta + \frac{\delta}{1-\delta} \, (1-2v)}{1+2v\eta + \frac{\delta}{1-\delta} \, (1-2v)} \right)}.$$

The first term in the RHS of (3.6) (see p. 26) therefore satisfies the inequality

$$S_k \Pr\Big(E_k(z_1), \underline{y} \in C_n(\theta)\Big) \leq \sqrt{\frac{1-2v}{1+2v\eta}} \; e^{-nE_2(c,v,\delta,\eta)} \tag{3.81}$$

where

$$\begin{cases} E_2(c,v,\delta,\eta) \triangleq -r_n(\delta) + \frac{1}{2}\ln(1-2v) + c \cdot \left(\frac{2v\eta + \frac{\delta}{1-\delta} \, (1-2v)}{1+2v\eta + \frac{\delta}{1-\delta} \, (1-2v)} \right) \\ \delta \triangleq \frac{k}{n}, \quad r_n(\delta) \triangleq \frac{\ln(S_k)}{n} \end{cases} . \tag{3.82}$$

The constraints on the exponents E_1 and E_2 in (3.69) and (3.82), respectively, are

$$s \geq 0, \quad v \leq 0, \quad 1 - 2s > 0, \quad 1 + 2v\eta > 0. \tag{3.83}$$

Rather than taking the derivative with respect to η in order to minimize the upper bound, for large n, Divsalar simply solved the equation

$$E_1(c, s, \eta) = E_2(c, v, \delta, \eta) \tag{3.84}$$

where the two exponents E_1 and E_2 are introduced in (3.68) and (3.81), respectively.

In order to simplify the calculations in the continuation of the analysis, a new set of four parameters (ρ, β, ξ and d) is introduced to replace the previous set of four parameters (s, v, η and δ):

$$\rho \triangleq \frac{s}{s - v}, \quad \beta \triangleq \rho(1 - 2v), \quad \xi \triangleq \frac{1}{1 + 2s\eta}, \quad d \triangleq \frac{\delta}{1 - \delta}. \tag{3.85}$$

From (3.83), the following constraints on the new set of parameters follow easily

$$0 \leq \rho \leq 1, \quad 0 \leq \beta \leq 1, \quad \frac{1}{1 + \eta} \leq \xi \leq 1, \quad 0 \leq d \leq \eta \tag{3.86}$$

where the constraint on d follows from the fact that the parameter δ (i.e., the normalized Hamming weight) appears only in the exponent E_2 in (3.82); this exponent corresponds to the case where $r_{z_1} \geq \beta_k(z_1)$, so that the inequality $d \leq \eta$ follows directly from (3.74). Armed with the new set of parameters in (3.85), it is straightforward to show that the exponents E_1 and E_2 in (3.69) and (3.82), respectively, are transformed to

$$E_1(c, \beta, \rho, \xi) = c - c\xi + \frac{1}{2} \ln\left(\frac{1 - \beta}{1 - \rho}\right) \tag{3.87}$$

and

$$E_2(c, \beta, \rho, \xi, d) = c - \frac{c\rho\xi}{(1 + \beta d)\xi - (1 - \rho)} + \frac{1}{2} \ln\left(\frac{\beta e^{-2r_n(\delta)}}{\rho}\right). \tag{3.88}$$

From (3.84), (3.87) and (3.88), it is straightforward to show that

$$c\xi - \frac{c\rho\xi}{(1+\beta d)\xi - (1-\rho)} + a = 0, \qquad a \triangleq \frac{1}{2}\ \ln\left(\frac{\beta(1-\rho)e^{-2r_n(\delta)}}{\rho(1-\beta)}\right). \tag{3.89}$$

By solving the quadratic equation in ξ, its value ξ (which should be non-negative) is

$$\xi = \frac{c-(1+\beta d)a + \sqrt{\big(c-(1+\beta d)a\big)^2 + 4c(1-\rho)(1+\beta d)a}}{2c(1+\beta d)} \tag{3.90}$$

This implies the coincidence of the two exponents in (3.87) and (3.88) (i.e., the requirement in (3.84) is fulfilled for this value of ξ).

The optimization of the parameter ρ is obtained by maximizing the exponent E_1 in (3.87) with respect to ρ, and showing that $\frac{\partial E_1}{\partial \rho} = 0$ if $a = 0$ (where a is introduced in (3.89)). This implies that the optimal value of ρ is given by

$$\rho = \frac{\beta e^{-2r_n(\delta)}}{1-\beta+\beta e^{-2r_n(\delta)}}\,. \tag{3.91}$$

Since the optimization with respect to ρ yields that $a = 0$, then we obtain from (3.90) that $\xi = \frac{1}{1+\beta d}$, and it follows from (3.85) and (3.87) that

$$E_1 = \frac{c\beta d}{1+\beta d} + \frac{1}{2}\ln\Big(1-\beta+\beta e^{-2r_n(\delta)}\Big), \quad d \triangleq \frac{\delta}{1-\delta}\,. \tag{3.92}$$

Finally, the maximization of the exponent E_1 with respect to β (which follows by solving the equation $\frac{\partial E_1}{\partial \beta} = 0$) gives

$$\begin{aligned}\beta &= \frac{1-\delta}{\delta}\left[\sqrt{\frac{c}{c_0(\delta)} + (1+c)^2 - 1} - (c+1)\right],\\ c_0(\delta) &\triangleq \Big(1-e^{-2r_n(\delta)}\Big)\left(\frac{1-\delta}{2\delta}\right).\end{aligned} \tag{3.93}$$

Surprisingly, this yields that *the error exponent of the Chernoff version of the TSB in* (3.92) *is equal to the error exponent of the Divsalar bound in* (3.41); this is true since also the optimized values of β in (3.39) and (3.93) coincide. This insightful result was made by Divsalar in [50, Section 4.D], and provides a closed form expression for the error exponent of the TSB. This error exponent is given by (3.92) with the parameters introduced in (3.93). For the error exponent in the limit where we let the block length tend to infinity, $r_n(\delta)$ is replaced by its asymptotic limit, i.e.,

$$r(\delta) \triangleq \lim_{n\to\infty} r_n(\delta).$$

For the ensemble of fully random binary linear block codes of rate R and block length n, the asymptotic growth rate of the distance spectrum is given by

$$r(\delta) = h(\delta) - (1-R)\ln 2$$

where h designates the binary entropy function to the natural base. Fig. 3.5 shows the error exponents associated with the random coding bound of Gallager [82] (to be presented in Section 4.2.1), the TSB and the union bound. It is clear from this figure that as the value of the code rate is increased, the gap between the error exponents of the TSB and the random coding bound of Gallager becomes more pronounced. The random coding error exponent of Gallager is positive for all the values of $\frac{E_{\mathrm{b}}}{N_0}$ above the value which corresponds to the channel capacity; according to the plots in Fig. 3.5, the error exponent of the TSB does not have this desired property of achieving capacity for fully random block codes.

3.2.10.2 Error exponent of a simplified tangential bound

In [50, Section 4.B], Divsalar derived the error exponent of the Chernoff version of the tangential bound. The error exponent of this simplified bound is of the form

$$E(c,\delta,\rho) = -\rho\, r_n(\delta) + \frac{\delta\rho c}{1-\delta+\delta\rho}\ , \quad c \triangleq \frac{E_{\mathrm{s}}}{N_0},\ \ 0 \le \rho \le 1 \qquad (3.94)$$

where δ is the normalized Hamming weight, and $r_n(\delta)$ is the normalized logarithm of the distance spectrum (see (3.82)). The optimal value of ρ which maximizes the error exponent is given by

$$\rho = \begin{cases} 1, & 0 \le \frac{r_n(\delta)}{c} \le \delta(1-\delta)\ , \\ \\ \sqrt{\frac{(1-\delta)c}{\delta\, r_n(\delta)}} - \frac{1-\delta}{\delta}, & \delta(1-\delta) \le \frac{r_n(\delta)}{c} \le \frac{\delta}{1-\delta}\ . \end{cases}$$

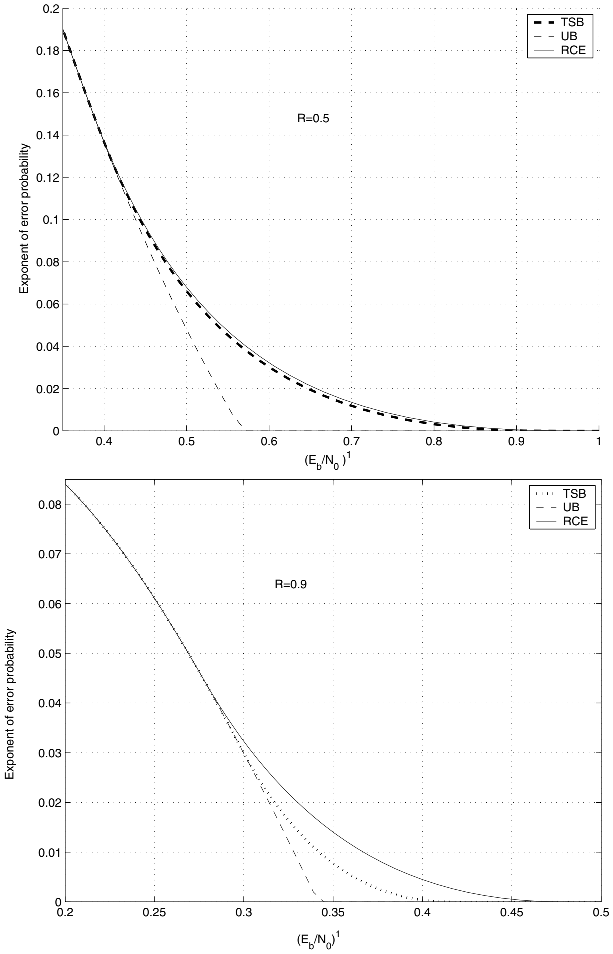

Fig. 3.5 Comparison between the error exponents for fully random block codes. These error exponents are based on the union bound (UB), the tangential-sphere bound (TSB) of Poltyrev [152] (which, according to [199], the error exponent of the TSB is identical to the error exponents of Divsalar's bound [50], the improved TSB and AHP bounds [224]), and the random coding bound (RCE) of Gallager [82] (see Section 4.2.1). The upper and lower plots refer to code rates of 0.5 and 0.9 bits per channel use, respectively, considering transmission over a binary-input AWGN channel. The error exponents are plotted versus the reciprocal of the energy per bit to the one-sided spectral noise density.

Finally, the substitution of the optimal value of ρ into (3.94) gives the error exponent of the simplified tangential bound

$$E(c,\delta) = \begin{cases} c\delta - r_n(\delta), & 0 \le \frac{r_n(\delta)}{c} \le \delta(1-\delta)\ , \\ \\ \left(\sqrt{c} - \sqrt{\frac{r_n(\delta)\,(1-\delta)}{\delta}}\right)^2, & \delta(1-\delta) \le \frac{r_n(\delta)}{c} \le \frac{\delta}{1-\delta}\ . \end{cases} \tag{3.95}$$

We note that the error exponent in (3.95) coincides with the error exponent of the Viterbi & Viterbi bound [209] (to be presented in Section 4.4.2).

In the asymptotic case where the block length goes to infinity, we obtain from (3.95) the following upper bound on the threshold under ML decoding

$$\left(\frac{E_\text{b}}{N_0}\right)_\text{threshold} \le \frac{1}{R} \max_{0<\delta<1} \frac{(1-\delta)\, r(\delta)}{\delta} \tag{3.96}$$

where $r(\delta) \triangleq \lim_{n\to\infty} r_n(\delta)$. Since $1 - e^{-x} < x$ for $x > 0$, we obtain that the upper bound given in (3.42) is tighter than the one in (3.96). Therefore, the common upper bound on the threshold which is provided by the simplified TSB and Divsalar's bound is tighter than the corresponding upper bound which follows from the simplified tangential bound.

3.2.10.3 Error exponent of a simplified sphere bound and comparison with Divsalar bound

Referring to Fig. 3.2, the sphere bound refers to the geometrical region $\mathcal{R}$ which is a sphere whose center is at $\gamma\underline{x}^0$. Based on the notation in Fig. 3.2, then $\eta = 1$ for the sphere bound, and the optimization is only with respect to the radius of the sphere $\mathcal{R}$. Since $\eta = 1$, then we obtain from (3.37) that $\xi = \beta$, and therefore the error exponent in (3.38) becomes

$$\begin{aligned} E(c,\delta,\beta,\rho) = -\rho r_n(\delta) - \frac{\rho}{2}\ln\left(\frac{\rho}{\beta}\right) \\ - \frac{1-\rho}{2}\ln\left(\frac{1-\rho}{1-\beta}\right) + \delta\beta c, \quad c \triangleq \frac{E_\text{s}}{N_0}. \end{aligned} \tag{3.97}$$

The maximization of the error exponent in (3.97) with respect to ρ gives the optimized value

$$\rho = \frac{1}{1 + \left(\frac{1-\beta}{\beta}\right) e^{2r_n(\delta)}}$$

and by substituting the optimized value of ρ in (3.97), we obtain the error exponent

$$E = -r_n(\delta) + \frac{1}{2} \ln\left(\beta + (1-\beta)e^{2r_n(\delta)}\right) + \delta\beta c. \tag{3.98}$$

Finally, the maximization of the error exponent in (3.98) gives the optimized value of β

$$\beta^* = \frac{2\delta c - \left[1 - e^{-2r_n(\delta)}\right]}{2\delta c\left[1 - e^{-2r_n(\delta)}\right]} \, .$$

We note that $\beta \geq 0$ in (3.37), and the error exponent in (3.98) vanishes at $\beta = 0$ and $\frac{\partial E}{\partial \beta} > 0$ for $0 < \beta < \beta^*$. Therefore, in order to have a positive error exponent, we need that $\beta^* > 0$, or equivalently

$$c > \max_{0<\delta<1} \frac{1 - e^{-2r_n(\delta)}}{2\delta} \, .$$

In the asymptotic case where the block length goes to infinity, we obtain an upper bound on the threshold under ML decoding which follows directly from the lower bound on c, and is given by

$$\left(\frac{E_{\mathrm{b}}}{N_0}\right)_{\text{threshold}} \leq \frac{1}{R} \max_{0<\delta<1} \frac{1 - e^{-2r(\delta)}}{2\delta} \tag{3.99}$$

where $r(\delta) \triangleq \lim_{n\to\infty} r_n(\delta)$. From the comparison of the upper bounds on the threshold values in (3.42) and (3.99), we see that the Divsalar bound yields a tighter upper bound on the threshold then the Chernoff version of the sphere bound (due to the factor of $1 - \delta$ which appears in the RHS of (3.42)). This improvement is attributed to the extra free parameter η in Fig. 3.2 which is optimized in the derivation of the Divsalar bound [50] (see Section 3.2.4), in contrast to the case with the sphere bound where the value of η is chosen *a-priori* to be 1.

3.2.11 Numerical results on the error probability of ML decoded binary linear block codes

We present in this section upper bounds on the decoding error probability of specific binary linear block codes and structured ensembles of turbo-like codes (e.g., regular LDPC codes [80, 81] and regular RA codes [54]).

For short block codes, we compare the upper bounds with the exact performance of these codes under ML decoding, so as to provide an indication about the tightness of the improved upper bounds, and the coding gain they provide as compared to union bounds (see Fig. 3.6). The performance of the $(24,12,8)$ Golay code under soft-decision ML decoding is estimated by computer simulations (i.e., the decoder decides on the codeword which corresponds to the largest correlation among the $2^{12} = 4096$ matched filters). The exact performance of the (63, 6, 32) bi-orthogonal code is calculated from (3.51), since it is known that the 1966 Viterbi bound holds in equality for binary linear block codes with fixed composition, such as orthogonal or bi-orthogonal codes (see [206, Section 8] and Section 3.2.7 here).

The upper and lower plots in Fig. 3.6 rely on the distance spectrum of the (24, 12, 8) Golay code and the (63, 6, 32) bi-orthogonal code, respectively. We note that the distance spectra of the binary (23, 12, 7) Golay code (which is a perfect code) and the extended (24, 12, 8) Golay code are well-known and given for example in [207, Table 2.2]. The (63, 6, 32) bi-orthogonal code includes $2^6 = 64$ codewords; except of the all-zero codeword, all the other 63 codewords of this code have a Hamming weight of 32. From the two plots in Fig. 3.6, one can see that the TSB bound is very tight, and it provides a significant improvement over the union bound. Although the error exponent of the Chernoff version of the TSB and the error exponent of the Divsalar bound coincide (see Section 3.2.10), it is obvious from Fig. 3.6 that especially for short codes, the TSB bound is significantly better than the Divsalar bound. Note that for these short block lengths, the tangential bound also outperforms the Divsalar bound, even though we have shown that the latter bound possess a superior exponent over the former bound.

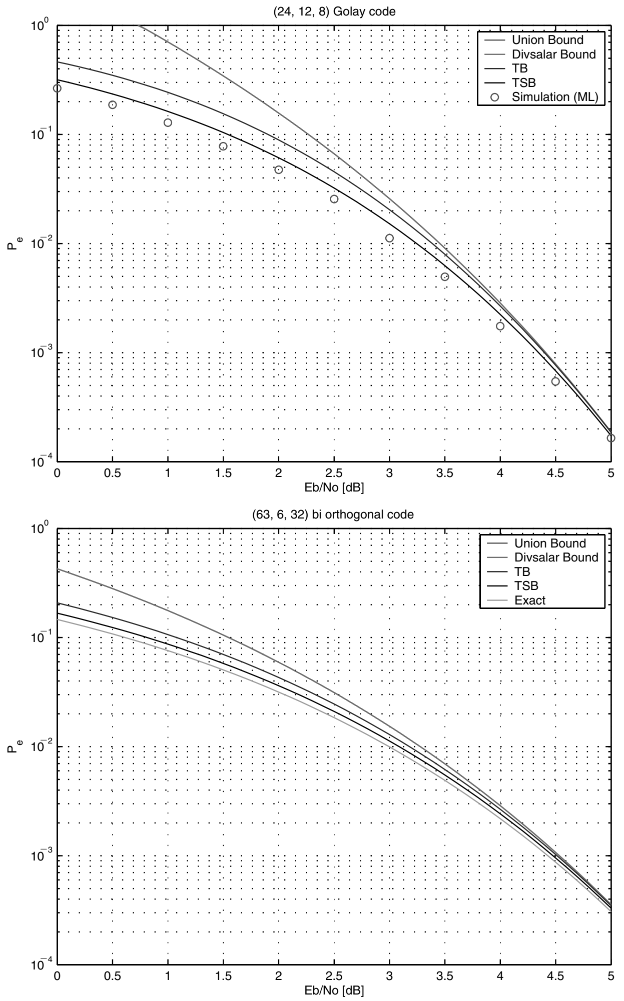

Fig. 3.6 Comparison between upper bounds on the decoding error probability of binary linear block codes with short length. The (24, 12, 8) Golay code and the (63, 6, 32) bi-orthogonal code are examined. Their transmission takes place over a binary-input AWGN channel, and the codes are ML decoded with soft-decision. The union bound, Divsalar's bound, the tangential bound and the tangential-sphere bound (TSB) are compared. The upper bounds on the block error probability for the two codes are compared with the exact decoding error probability. For the (24, 12, 8) Golay code, we obtain simulation results of a soft-decision ML decoder. For the bi-orthogonal code, the exact error probability is calculated from Eqs. (3.51) and (3.52) (see Section 3.2.7). The union and Divsalar bounds coincide in these two plots.

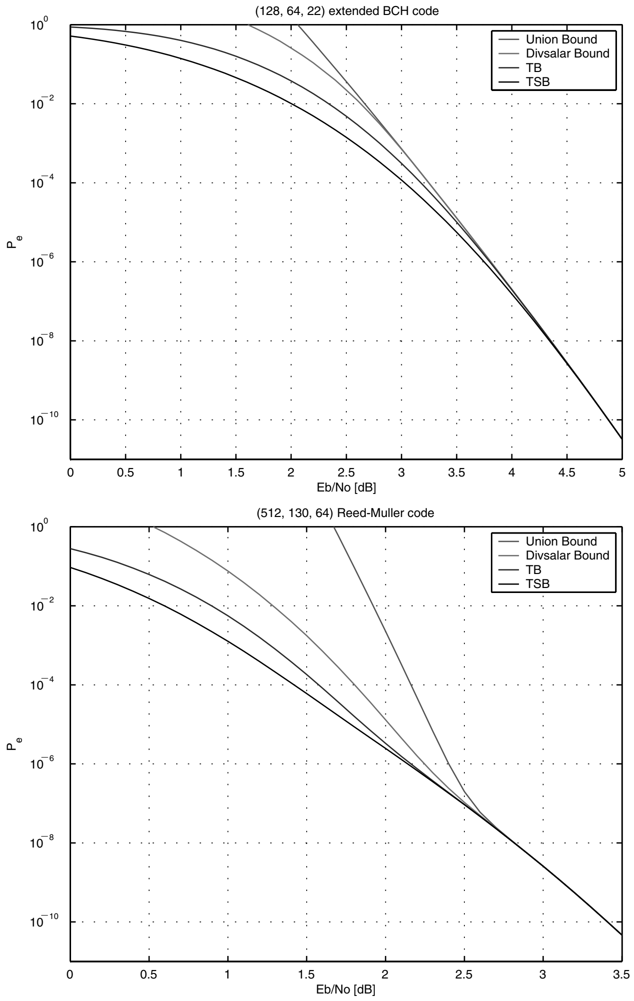

Fig. 3.7 Comparison between upper bounds on the decoding error probability of binary linear block codes with moderate length. The (128, 64, 22) extended BCH code (upper plot) and the third order (512, 130, 64) Reed-Muller code (lower plot) are examined. Their transmission takes place over a binary-input AWGN channel, and the codes are ML decoded with soft-decision. The union bound, Divsalar's bound, tangential bound and the tangential-sphere bound (TSB) are compared.

Upper bounds on the decoding error probability of the extended (128, 64, 22) BCH code and the third-order (512, 130, 64) Reed-Muller code are depicted in Fig. 3.7; it is assumed that the codes are BPSK modulated, transmitted over the AWGN channel, coherently detected and ML decoded. The plots in Fig. 3.7 rely on the distance spectra of the extended (128, 64, 22) BCH code and the third-order (512, 130, 64) Reed-Muller code which are provided in [46] and [190], respectively (see also [112]).

Upper bounds on the performance of ML decoding for some ensembles of (n, j, k) LDPC codes with rate one-half are presented in Fig. 3.8. The effect of the block length n on the gap to the Shannon capacity limit is also demonstrated. The error bounds are based on Gallager's upper bound on the ensemble distance spectrum of regular LDPC codes (see [81, Theorem 2.3]). Based on the upper sub-matrices of the parity-check matrices of Gallager's ensembles (see [81, Fig. 2.1]), it follows directly that the Hamming distance of all codewords of these ensembles should be *even*; the average distance spectra of these ensembles is therefore equal to zero for all the Hamming weights which are odd. In order to obtain upper bounds on the performance of these ensembles, Gallager's upper bound on the distance spectrum is combined here with two possible upper bounds on the ML decoding error probability: the TSB, and the union bound in its Q-form. It is also evident from the lower plot in Fig. 3.8 that for large block lengths, the union bounds are essentially useless at rates above the cutoff rate (for a binary-input AWGN channel and a transmission rate of one-half bit per channel use, the cutoff rate corresponds to $\frac{E_\mathrm{b}}{N_0} = 2.45$ dB). In both plots of Fig. 3.8, the same values of j and k (i.e., the degree of the variable nodes and the degree of the parity-check nodes, respectively) are compared for ensembles of (n, j, k) LDPC codes with block lengths of $n = 1$ K and 10 K coded bits; it reflects the significant improvement that results by increasing n by a factor of 10. For example, for the case of $j = 6$, $k = 12$, the values of $\frac{E_\mathrm{b}}{N_0}$ required by the TSB for achieving a *block error probability* of 10^{-4} are 1.80 dB and 0.73 dB for $n = 1$ K and 10 K bits, respectively (while the Shannon capacity limit corresponds in this case to 0.19 dB). The considerable advantage of the tangential sphere bounding technique over the union bounds (especially for large block lengths) is demonstrated in Fig. 3.8 (see also [168, 170]). For example, the gain achieved for a block error probability of 10^{-5} is 1.72 dB in the case that $n = 10$ K bits, $j = 6$ and $k = 12$. As exhibited in Fig. 3.8, there is only a slight improvement in the

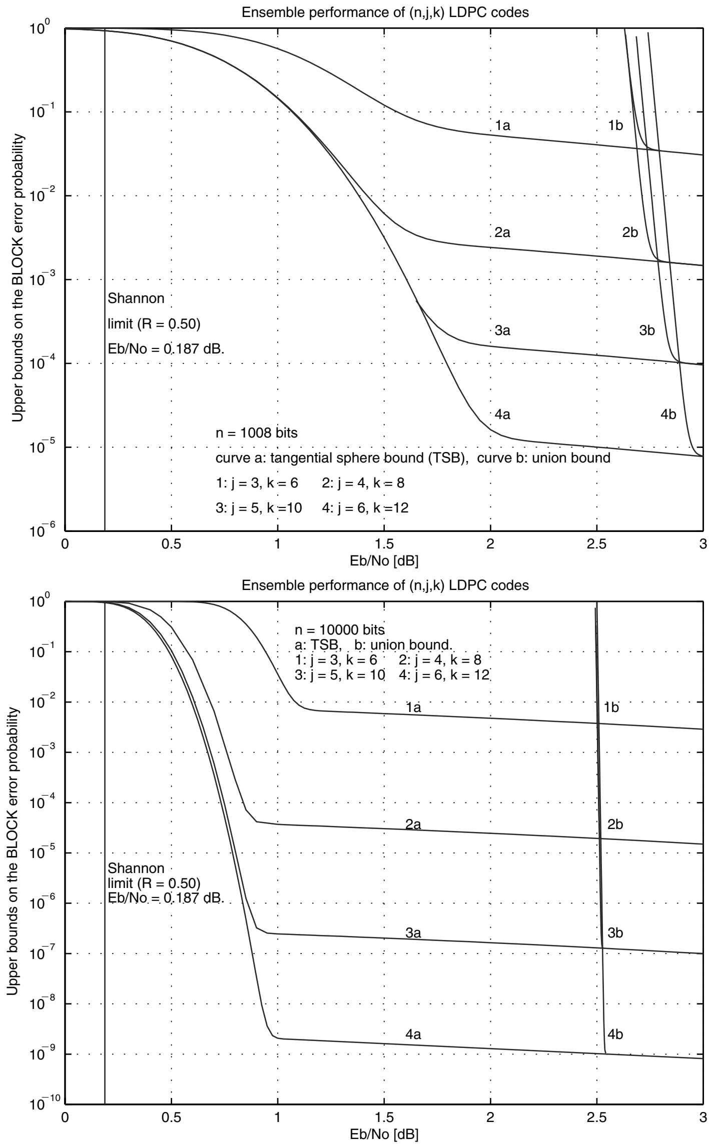

Fig. 3.8 Upper bounds on the block error probability of ML decoding for ensembles of (n, j, k) regular LDPC codes in a binary-input AWGN channel where $n = 1008$ (upper plot) and $n = 10000$ (lower plot), $j = 3, 4, 5, 6$ and $k = 2j$ (rate of one half). The upper bounds are based on Gallager's upper bound on the ensemble distance spectrum of (n, j, k) LDPC codes and the TSB. Union bounds in Q-form appear for comparison.

ensemble performance of (n,j,k) LDPC codes, by increasing the value of j above 6 (while keeping $k=2j$ for maintaining the code rate fixed). Therefore, the ML block error probabilities of ensembles of (n,j,k) LDPC codes, are investigated here for $j=6$ for a variety of rates and block lengths of $n=5$ K, 10 K, 20 K and 40 K bits. The results of this comparison are summarized in Table 3.1 (see p. 77), demonstrating the impressive potential performance of ensembles of LDPC codes of length in the range 5 K–40 K bits. This table is created by comparing an upper bound on the value of required $\frac{E_{\text{b}}}{N_0}$ for achieving a block error probability of 10^{-5} by ML decoding with a lower bound on the Shannon capacity (as R is lower bounded by $1-\frac{j}{k}$ [81], the value of $\frac{E_{\text{b}}}{N_0}$ that corresponds to the channel capacity with this lower bound on the code rate, is actually a lower bound on the Shannon capacity that corresponds to the exact code rate. However, for large block length n, this lower bound on the rate is very tight). Fig. 3.9 presents results for ensembles of (n,j,k) LDPC codes of rate 0.250. For example in the case of $j=6$, $k=8$ and $n=40$ K, an upper bound on the value of $\frac{E_{\text{b}}}{N_0}$ required to achieve a block error probability of 10^{-5} with ML decoding is -0.46 dB, that is only 0.33 dB away from the channel capacity (see also Table 3.1).

The ensemble performance of uniformly interleaved repeat-accumulate (RA) codes was considered by Divsalar, Jin and McEliece [54]. The structure of these codes and the expressions for the input-output weight enumerator function (IOWEF) of such ensembles of codes is introduced in [54] (see the upper plot in Fig. 2.1). The average number of codewords in the ensemble of uniformly interleaved (qN,N) RA codes having an information weight w and an overall Hamming weight ℓ is

$$A_{w,\ell}^{(N)} = \frac{\binom{N}{w}\binom{qN-\ell}{\lfloor qw/2\rfloor}\binom{\ell-1}{\lceil qw/2\rceil-1}}{\binom{qN}{qw}} \tag{3.100}$$

where $0\le w\le N$, $0\le \ell\le qN$, and $\lfloor x\rfloor, \lceil x\rceil$ denote the maximal and minimal integers, respectively, satisfying the inequality $\lfloor x\rfloor \le x\le \lceil x\rceil$.

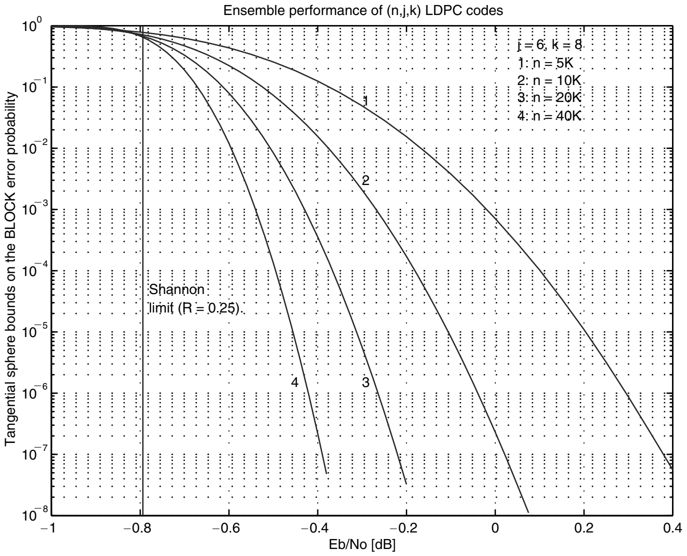

Fig. 3.9 Upper bounds on the block error probability of some ensembles of (n,j,k) regular LDPC codes whose rate is $\frac{1}{4}$ bit per channel use (where $j = 6$ and $k = 8$). The transmission of the codes takes place over a binary-input AWGN channel, and the bounds refer to ML decoding. The upper bounds are based on Gallager's upper bound on their ensemble distance spectrum [81, Section 2.2] and the TSB of Poltyrev. The figure refers to block lengths of $n = 5$ K, 10 K, 20 K and 40 K bits.

Based on the IOWEF of uniformly interleaved RA codes in (3.100), the distance spectra of some ensembles of (qN, N) RA codes, are illustrated in the upper plot of Fig. 3.10, versus the normalized Hamming weight of the codewords (the normalization is with respect to the block length qN), where $N = 1024$ and $q = 2, 3, 4$. It is observed in the upper plot of Fig. 3.10 that for a fixed value of N, the number of codewords with relatively low Hamming weights is reduced by increasing the value of q (note that the overall number of codewords (2^N) remains constant). Therefore, it is reasonable (as indicated in the lower plot of Fig. 3.10) that the ensemble performance of the uniformly interleaved and serially concatenated RA codes is improved by increasing the value of q; by increasing the value of q, the interleaver length (qN) is increased which enhances the interleaver gain of such an ensemble of interleaved and serially concatenated codes, but also decreases the code rate. Upper bounds on the block and bit error probability of uniformly interleaved

Table 3.1 The value of energy per bit to noise spectral density $\left(\frac{E_b}{N_0}\right)$ required for an upper bound on the block error probability of 10^{-5} with soft-decision ML decoding for the Gallager ensemble of (n,j,k) regular LDPC codes with $j = 6$. The bounds are based on the TSB. The gaps (in dB) between the values of $\frac{E_b}{N_0}$ achieving an upper bound on the block error probability of 10^{-5} to the Shannon capacity of a binary-input AWGN channel are written in parenthesis.

The number of ones (k) in each row of the parity matrix H of the ensemble of codes and a tight lower bound on the rate (R).	$k = 8$ $R = 0.250$	$k = 12$ $R = 0.500$	$k = 24$ $R = 0.750$
The block length n	The value of $\frac{E_b}{N_0}$ (and the gap to capacity)		
$n = 5$ K bits	0.20 dB (0.99 dB)	1.00 dB (0.81 dB)	2.57 dB (0.94 dB)
$n = 10$ K bits	−0.11 dB (0.68 dB)	0.79 dB (0.60 dB)	2.39 dB (0.76 dB)
$n = 20$ K bits	−0.31 dB (0.48 dB)	0.64 dB (0.45 dB)	2.26 dB (0.63 dB)
$n = 40$ K bits	−0.46 dB (0.33 dB)	0.54 dB (0.35 dB)	2.18 dB (0.55 dB)

RA codes under soft-decision ML decoding are illustrated in the lower plot of Fig. 3.10. The usefulness of the TSB is demonstrated at rates considerably above the channel cutoff rate: for $q = 3,4$ (the code rate of the RA code is $\frac{1}{3}$ or $\frac{1}{4}$, respectively), the cutoff rate of the binary-input AWGN channel corresponds to $\frac{E_b}{N_0} = 2.03$ dB and 1.85 dB, respectively. From the lower plot of Fig. 3.10, we obtain that for $q = 3$, the TSB on the bit error probability [170] is equal to 10^{-5} at $\frac{E_b}{N_0} = 1.58$ dB (i.e., 0.45 dB below the value of $\frac{E_b}{N_0}$ that corresponds to the channel cutoff rate). For $q = 4$, the TSB is equal to 10^{-5} at $\frac{E_b}{N_0} = 0.79$ dB (i.e., 1.06 dB below the value of $\frac{E_b}{N_0}$ which corresponds to the cutoff rate, and 1.59 dB above the value of $\frac{E_b}{N_0}$ which corresponds to the channel capacity). It is also shown in the lower plot of Fig. 3.10 that the union bounds in Q-form are useless at rates beyond the cutoff rate of the channel, as expected for long enough block codes (since the interleaver length for these RA codes is 3072 and 4096 for $q = 3,4$, respectively).

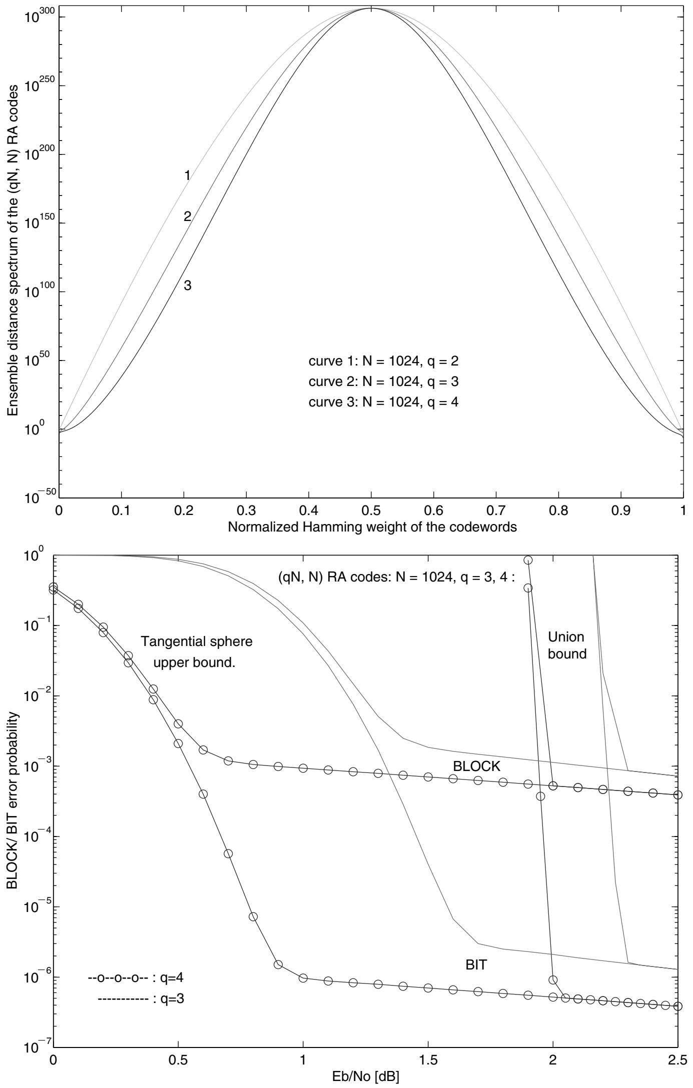

Fig. 3.10 Distance spectra and performance bounds of uniformly interleaved RA codes [54] under ML decoding (see the upper plot of Fig. 2.1 on p. 19). The upper plot shows the distance spectra of some ensembles of (qN, N) RA codes versus the normalized Hamming weights of their codewords (the normalization is with respect to the block length qN) where $N = 1024$ and $q = 2, 3, 4$. The lower plot shows upper bounds on the block and bit error probabilities of the two ensembles with $q = 3, 4$. The considered upper bounds are the TSB and the union bound in Q-form.

3.3 Improved upper bounds for fading channels

In this section, we present rigorous analytical upper bounds on the ML decoding error probability of binary linear block codes, operating over fading channels (for a tutorial paper on fading channels, the reader is referred to [26]). We mainly focus in this section on fully interleaved (memoryless) Rician fading channels. These bounds are applied to several ensembles of turbo-like codes, demonstrating their significant advantage over the union bounds in a portion of the rate region above the channel cutoff rate (R_0). Throughout this section, we assume a perfect side information on the channel i.i.d. fading samples which are also available to the receiver. The theoretical implications of this assumption are discussed in [114]. The model of block-fading channels and related bounds are shortly addressed at the end of this section.

3.3.1 Fully interleaved fading channels: System model, channel capacity and cutoff rate, and union bounds

3.3.1.1 The system model

The model of the communication system is the following: the information bits are encoded, fully interleaved and BPSK modulated. The modulated signal is transmitted through a frequency non-selective fading channel. We discuss in this section fully interleaved Rician fading channels and also Rayleigh fading channels where the latter is combined with space diversity of order L, based on the maximum ratio combining (MRC) principle. As a consequence of a perfect channel interleaver (which clearly differs from the interleaver of the code), the fading samples which correspond to the interleaved coded bits are assumed to be i.i.d. The noise is an AWGN with a zero mean and a double-side spectral density of N_0. At the receiver, assumed to be equipped with perfect channel state information (CSI) of the realizations of the fading values, we assume that the fading samples received by these L antennas are statistically independent. The received signal is coherently detected. Finally, the demodulated bits are deinterleaved (according to the inverse permutation of the channel interleaver) and ML decoded.

Considering a turbo code, we assume a termination to the all-zero state at the end of each frame (block). Clearly, no termination is required for block codes in general and for LDPC codes and RA codes in particular.

3.3.1.2 The capacity and cutoff rate of binary-input fully interleaved fading channels

We denote the pdf of the non-negative[4] fading a by $p(a)$. Clearly, with a (ideally given to the receiver) interpreted as part of the measurements and independent of the transmitted signals, then

$$\begin{array}{l} p_0(y,a) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(y - a\sqrt{\frac{2E_\mathrm{s}}{N_0}}\right)^2\right] \cdot p(a) \\ p_1(y,a) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(y + a\sqrt{\frac{2E_\mathrm{s}}{N_0}}\right)^2\right] \cdot p(a) \end{array}, \qquad \begin{array}{c} -\infty < y < \infty \\ a \geq 0 \end{array} \tag{3.101}$$

where $\frac{E_\mathrm{s}}{N_0}$ stands for the energy per symbol to the spectral noise density.

For MBIOS channels in (3.101), the capacity-achieving distribution is clearly symmetric: $\Pr(x=0) = \Pr(x=1) = \frac{1}{2}$. A straightforward calculation of the channel capacity (C) of the fully interleaved fading channels in (3.101) gives that

$$C = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \int_0^{\infty} p(a) \exp\left[-\frac{1}{2}\left(y + a\sqrt{\frac{2E_\mathrm{s}}{N_0}}\right)^2\right] \cdot \log_2\left(\frac{2}{1 + \exp\left(2ay\sqrt{\frac{2E_\mathrm{s}}{N_0}}\right)}\right) da\, dy \tag{3.102}$$

in units of bits per channel use.

[4] The effect of the phases of the fading measurements is eliminated at the receiver and the fades during each symbol are treated as non-negative random variables.

For a fully interleaved Rician fading channel, the probability density function of the amplitudes of the i.i.d. fading samples is

$$p(a) = \begin{cases} 2a(K+1)\exp(-(K+1)a^2 - K)\, I_0\left(2\sqrt{K(K+1)}\,a\right), & a \geq 0 \\ 0 & a < 0 \end{cases} \tag{3.103}$$

where the Rician parameter K stands for the power ratio of the direct to the diffused received path of the fading channel, and $I_0(x)$ designates the Modified Bessel function of order zero:

$$I_0(x) = \frac{1}{\pi}\int_0^{\pi} \exp(x\cos\theta)\, d\theta\ , \quad x \in \mathbb{R}.$$

The probability density function of the i.i.d. fading samples of a fully interleaved Rayleigh fading channel with MRC diversity of order L admits the form

$$p(a) = \begin{cases} \dfrac{2L^L a^{2L-1}\exp(-La^2)}{(L-1)!} & a \geq 0 \\ 0 & a < 0 \end{cases} .$$

Clearly, for coded communications: $E_{\mathrm{s}} = RE_{\mathrm{b}}$ where E_{s} and E_{b} designate the energies per coded symbol and per information bit, respectively, and R is the rate of the code (in bits per channel use). The value of $\frac{E_{\mathrm{b}}}{N_0}$ which corresponds to the channel capacity for a certain code rate R is calculated numerically by solving the implicit equation $C = R$, where C is expressed in (3.102), associated with the appropriate expression for the probability density function $p(a)$ of the i.i.d. fading samples.

When the bit error probability is restricted not to exceed a certain value P_{b}, then by applying the rate-distortion theory with the Hamming distance between two binary codewords as a distortion measure, the minimal theoretical value of $\frac{E_{\mathrm{b}}}{N_0}$ is found by solving numerically the implicit equation $C = R\left(1 - h(P_{\mathrm{b}})\right)$ where $h(x)$ designates the binary entropy function to the base 2, and the capacity C of general fully interleaved fading channels is expressed in (3.102), assuming perfect CSI of the i.i.d. fading samples at the receiver.

The cutoff rate (in bits per channel use) of a fully interleaved Rayleigh fading channel with MRC space diversity of order L admits the form

$$R_0 = 1 - \log_2 \left(1 + \left(1 + \frac{E_{\mathrm{s}}}{LN_0} \right)^{-L} \right)$$

or alternatively, the value of $\frac{E_{\mathrm{b}}}{N_0}$ which corresponds to the cutoff rate is expressed in terms of the code rate R and the order of diversity L by the equality

$$\frac{E_{\mathrm{b}}}{N_0} = \frac{L}{R} \left[(2^{1-R} - 1)^{-\frac{1}{L}} - 1 \right] . \tag{3.104}$$

For a fully interleaved Rician fading channel, the cutoff rate is given by

$$R_0 = 1 - \log_2(1 + z) \tag{3.105}$$

bits per channel use, where z stands for the Bhattacharyya constant of this channel. By conditioning on the magnitude of the fading a, the value of the Bhattacharyya constant (2.3) is $\exp(-\frac{a^2 E_{\mathrm{s}}}{N_0})$, and averaging with respect to a gives

$$z = \int_0^{\infty} \exp(-\frac{a^2 E_{\mathrm{s}}}{N_0})\, p(a)\, \mathrm{d}a$$

where $p(a)$ is given by (3.103). An alternative expression for this integral is given in [172, Eq. (8)].

3.3.1.3 The pairwise error probability and union bounds

For a fully interleaved Rician fading channel with a Rician factor K, the expression of the pairwise ML decoding error probability for two codewords which differ in d symbols was derived in [188]. This derivation relied on Craig's identity for the Q-function in (2.9).

Under the assumptions in Section 3.3.1.1, the following expressions for the pairwise error probability refer to the case where the Hamming distance between the transmitted codeword ($\underline{x}$) and another competitive codeword ($\hat{\underline{x}}$) is d. For a fully interleaved Rician fading channel whose Rician factor is equal to K, the expression for the pairwise error probability follows from combining [188, Eqs. (25), (30)]:

$$\begin{aligned} &\Pr\left(\underline{x} \to \hat{\underline{x}} \mid W_H(\underline{x}, \hat{\underline{x}}) = d\right) \\ &\quad = \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \left(\frac{1+K}{1+K+\frac{E_{\mathrm{s}}/N_0}{\sin^2\theta}} \right)^d e^{-\frac{Kd}{1+\frac{(1+K)\sin^2\theta}{Es/N_0}}}\, d\theta \end{aligned} \tag{3.106}$$

where $W_H(\underline{x}, \hat{\underline{x}})$ designates the Hamming distance between the binary codewords $\underline{x}$ and $\hat{\underline{x}}$. We note that an alternative expression for the pairwise error probability of fully interleaved Rayleigh fading channels with perfect CSI is given in [101] (hence, it is an alternative expression to (3.106) for the particular case where $K = 0$).

For fully interleaved Rayleigh fading channels with MRC space diversity of order L (where the fading samples received by the L antennas are assumed to be i.i.d.), the expression for the pairwise ML decoding error probability results in by substituting $K = 0$ in (3.106), and by also replacing E_s and d in the left hand side of (3.106) with $\frac{E_\mathrm{s}}{L}$ and dL, respectively:

$$\Pr\left(\underline{x} \rightarrow \hat{\underline{x}} \mid W_\mathrm{H}(\underline{x}, \hat{\underline{x}}) = d\right) = \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \left(\frac{1}{1 + \dfrac{E_\mathrm{s}/N_0}{L \sin^2 \theta}} \right)^{dL} d\theta \, . \quad (3.107)$$

The union bound on the block error probability for an ML decoded linear and binary block code admits the form

$$P_\mathrm{e} \leq \sum_{d=d_\mathrm{min}}^{n} S_d \Pr\Big(\underline{x} \rightarrow \hat{\underline{x}} \mid W_\mathrm{H}(\underline{x}, \hat{\underline{x}}) = d\Big) \quad (3.108)$$

where d_min denotes the minimal Hamming weight of the nonzero codewords of the code $\mathcal{C}$, and the expressions of the pairwise error probability for the fully interleaved fading channels are provided in (3.106) and (3.107). Here, S_d is the appropriate coefficient of the distance spectrum of the code, which designates the number of codewords whose Hamming weight equals d.

Similarly, the union bound on the bit error probability admits the form

$$P_\mathrm{b} \leq \sum_{d=d_\mathrm{min}}^{n} S'_d \Pr\Big(\underline{x} \rightarrow \hat{\underline{x}} \mid W_\mathrm{H}(\underline{x}, \hat{\underline{x}}) = d\Big) \quad (3.109)$$

where S'_d is introduced in (3.13).

Union bounds on the ML decoding error probability of binary linear codes whose transmission takes place over fully interleaved fading channels with perfect CSI at the receiver are considered in [5, 89, 101]; these bounds are exemplified there for various ensembles of turbo codes.

3.3.2 Generalization of the Engdahl and Zigangirov bound for fully interleaved fading channels

The Engdahl and Zigangirov bound in [67] provides an upper bound on the ML decoding error probability when the communications takes place over the AWGN channel (see Section 3.2.6). Sason and Shamai generalized the Engdahl and Zigangirov bound for fully interleaved fading channels [172, Section 3.B]. In the following, we present the guidelines for the derivation of the generalized bound. For further details, we refer the reader to [172, Section 3.B and Appendices B and C].

Let's assume a binary linear block code $\mathcal{C}$ of length n and dimension k which is BPSK modulated, and transmitted over a fully interleaved fading channel. As noted in Section 3.3.1, we assume here that a perfect channel side information (CSI) is available at the receiver. Let the vector $\underline{a} = (a_1, a_2, \ldots, a_n)$ denote the perfect measurements of the i.i.d. fading samples, where $E\left[a_i^2\right] = 1$. We assume that the symbols 0 and 1 are mapped by the BPSK modulator to $+\sqrt{E_\mathrm{s}}$ and $-\sqrt{E_\mathrm{s}}$, respectively. As before, we also assume without any loss of generality that the all-zero codeword is transmitted (due to the linearity of the code and the symmetry of the channel). The components of the demodulated signal ($\underline{y}$) are

$$y_i = a_i \sqrt{E_\mathrm{s}} + \nu_i \,, \quad i = 1, 2, \ldots, n$$

where $\{a_i\}_{i=1}^n$ are the i.i.d fading samples, and $\{\nu_i\}_{i=1}^n$ are the i.i.d. zero-mean Gaussian noise samples with variance $E[\nu_i^2] = \frac{N_0}{2}$. Let us define the random variable (RV)

$$z \triangleq \sum_{i=1}^{n} a_i \, y_i = \sum_{i=1}^{n} \Big(a_i^2 \sqrt{E_\mathrm{s}} + a_i \, \nu_i \Big). \tag{3.110}$$

We now partition the binary and linear block code $\mathcal{C}$ into constant Hamming weight subcodes ($\mathcal{C}_d$), where the subcode $\mathcal{C}_d$ ($d = 1, 2, \ldots, n$) includes all the codewords of $\mathcal{C}$ with a Hamming weight d and also includes the all-zero codeword.

The fading samples of the channel are i.i.d. RVs, so the pairwise ML decoding error probability doesn't change for all the competitive codewords of the same Hamming weight (d). Since we consider an MBIOS channel, then we assume without any loss of generality that a competitive codeword of Hamming weight d differs from the transmitted all-zero codeword in its first d coordinates, i.e, the first d coordinated of the competitive codeword are ones and the succeeding $n-d$ components are zeros.

Let W_d (where $d = 1,2\dots n$) be the RV

$$W_d \triangleq \sum_{i=1}^{d} a_i\, y_i = \sum_{i=1}^{d} \Big(a_i^2\, \sqrt{E_{\rm s}} + a_i\, \nu_i\Big).$$

If $W_d < 0$, then such a codeword of Hamming weight d is preferred by the ML decoder (rather than the all-zero codeword which was transmitted).

The case where z is small enough may yield an unreliable decision of the ML decoder (it may happen, for example, when the channel undergoes a severe fade). We therefore introduce a threshold η_z, to be optimized in order to derive the tightest upper bound among the following family of upper bounds on the ML decoding error probability:

$$P_{\rm e} \le \sum_{d=1}^{n} \Big\{S_d \cdot \Pr\Big(W_d < 0,\, z \ge \eta_z\Big)\Big\} + \Pr\Big(z < \eta_z\Big) \qquad (3.111)$$

where $\{S_d\}_{d=0}^{n}$ is the distance spectrum of the block code $\mathcal{C}$. We note that (3.111) follows as a particular case of the general inequality in (3.1) where the region $\mathcal{R}$ in the latter inequality is chosen to be a half-space (see (3.110)), and where we apply the union bound to the first term in the RHS of (3.1). We also note that in the limit where $\eta_z \to -\infty$, the upper bound in (3.111) turns to be the well known union bound in (3.108). Therefore, an optimization of the RHS of (3.111) over the parameter η_z yields an upper bound which is uniformly tighter than the union bound, and also cannot exceed unity (since in the limit where $\eta_z \to \infty$, we get the trivial bound $P_{\rm e} \le 1$. The final version of this bound is expressed explicitly in [172, Eq. (38)] (with some algebraic simplifications in [172, Appendix C], and the optimized free parameter of this bound is given implicitly as a solution to the optimization equation [172, Eq. (39)]. Note that the existence and uniqueness of a solution to the latter equation for the optimized free parameter in (3.111) is proved in [172, Appendix B]. In particular, it is shown in [172, Section III.B] that the generalization of the Engdahl and Zigangirov bound for fully interleaved fading channels particularizes to the Engdahl and Zigangirov bound in [67] (see Section 3.2.6) for the case of a binary-input AWGN channel.

An upper bound on the *bit error probability* for an ensemble of linear, binary block codes results in by simply replacing the average distance spectrum $\{S_d\}_{d=0}^n$ of the ensemble of codes by the sequence $\{S'_d\}_{d=0}^n$, as defined in (3.13).

The generalization of the Engdahl and Zigangirov bound is applied in the sequel to ensembles of turbo-like codes whose transmission takes place over fully interleaved fading channels, and the tightness of this bound is compared to other reported bounds (see Section 4.7.2).

In the block-fading channel model, an n-length codeword is affected by a number (M) of independent fading gains. In this case, an n-length codeword is split into M blocks where M divides n; over each block, the channel fading is so highly correlated that it is considered to be constant. The Engdahl and Zigangirov bound was recently generalized for the performance analysis under ML decoding of block codes whose transmission takes place over block-fading channels. For further details, the reader is referred to [218]. For small values of M, the generalization of the Engdahl and Zigangirov bound for block-fading channels shows an improvement over previously reported bounds (see, e.g., the Malkamaki-Leib bound in [126]).

3.3.3 Generalization of the Viterbi and Viterbi bound

The Viterbi and Viterbi bound is an upper bound on the ML decoding error probability of binary linear block codes operating over a binary-input AWGN channel [209]. A generalization of this bound for fully interleaved Rician fading channels with perfect CSI at the receiver was derived in [173] (see Section 2 and the Appendix in [173] where the generalization of the Viterbi and Viterbi bound for fully interleaved Rician fading channels was derived in two alternative ways). The generalized bound admits the form

$$P_{\mathrm{e}} \leq \sum_d P_{\mathrm{e}}(d)$$

where

$$P_{\mathrm{e}}(d) \le (S_d)^{\rho} \left(\frac{1+K}{1+K+\beta_1}\right)^{d\rho} \cdot \exp\left(-\frac{K\beta_1\, d\rho}{1+K+\beta_1}\right)$$
$$\cdot \left(\frac{1+K}{1+K+\beta_2}\right)^{(n-d)\rho} \cdot \exp\left(-\frac{K\beta_2(n-d)\rho}{1+K+\beta_2}\right)$$
$$\cdot \left(\frac{1+K}{1+K+\beta_3}\right)^{n(1-\rho)} \cdot \exp\left(-\frac{K\beta_3 n(1-\rho)}{1+K+\beta_3}\right)$$

and

$$\beta_1 = \frac{E_{\mathrm{s}}}{N_0}, \quad \beta_2 = \frac{E_{\mathrm{s}}}{N_0} \cdot \left[1 - \Big(1+\zeta(1-\rho)\Big)^2\right],$$
$$\beta_3 = \frac{E_{\mathrm{s}}}{N_0} \cdot \left[1 - (1-\zeta\rho)^2\right].$$

The optimization of the bound is performed with respect to the two parameters ρ and ζ where $0 \le \rho \le 1$, and $1 + K + \beta_i > 0$ for $i = 1, 2, 3$.

For a binary-input AWGN channel, since $K \to \infty$, it can be verified (see [173]) that the generalized bound above reduces to the Viterbi and Viterbi bound in [209] for the optimal value $\zeta = \frac{\delta}{1-\delta+\delta\rho}$ where $\delta \triangleq \frac{d}{n}$ is the normalized weight of codewords with Hamming weight d and block length n.

We note that the generalization of the Viterbi and Viterbi bound for fully interleaved fading channels with perfect CSI, as reported in [11, 10], seems to be problematic. Our reservation stems from the fact that the Viterbi and Viterbi bounding technique in [209] is invalidated once the parameter $H = \exp(-\frac{E_{\mathrm{s}}}{N_0})$ (which is the Bhattacharyya constant for a binary-input AWGN channel) is replaced by the one corresponding to fast Rician fading channels (see [11, Eq. (12)]). The specific correlations in the derivation of the bound in [209] demand special care when generalized to fading channels. The problem with the generalization in [11, 10] is that in contrast to the particular case of a binary-input AWGN channel, for fully interleaved fading channels, we obtain that the covariance between the outputs of the two correlators which correspond to the all-zero codeword and to an arbitrary codeword of Hamming weight d does not depend only on d (i.e., as a result of the i.i.d. fading samples, this covariance does not stay constant with respect to all the codewords of Hamming weight d). A generalization of the Viterbi and Viterbi bound for fully interleaved fading channels with imperfect CSI at the receiver is considered in [186]. Again, the derivation in [186] is problematic since it relies on the arguments of [11, 10], when now H is replaced by the one corresponding to imperfect CSI.

We note that the generalized Viterbi and Viterbi bound for fully interleaved Rician fading channels is derived in [173, Section 2] as a particular case of the Gallager-type bounds; these bounds are discussed extensively in the next section.

Bounds on the ML decoding error probability for fully interleaved fading channels which are presented in this section, and some variations of the Gallager bounds which are particularized to fading channels are applied to various ensembles of codes in Section 4.7.2.

3.3.4 Improved bounds for block fading channels

The block-fading channel model is introduced for modelling slowly-varying fading. It is particularly relevant in wireless communications where slow time-frequency hopping or multi-carrier modulation using orthogonal frequency division multiplexing (OFDM) are considered. This model is a particular case of parallel channels which are considered in the next section (see Sections 4.6 and 4.7.5). For various upper bounds on the ML decoding error probability of linear codes whose transmission takes place over block-fading channels, the reader is referred to [9, 27, 70, 96, 109, 119, 122, 126, 219, 218, 227, 228] and references therein.

3.4 Concluding comments

We introduce in this section rigorous upper bounds on the block and bit error probabilities of ML decoded linear codes. The performance analysis here refers to Gaussian and fading channels.

For modulated signals of constant energy whose transmission takes place over an AWGN channel, the tangential-sphere bound of Poltyrev [152] happens to be one of the tightest known upper bounds on the ML decoding error probability; however, this bound fails to reproduce the random coding error exponent for ensembles of fully random block codes.

Performance bounds are reviewed in this section for both fully interleaved and block-fading channels, and other bounds are also considered in the following section. The reader is referred to [78, 105, 116] which address performance bounds for correlated fading channels.

All the improved upper bounds introduced in this section are not subject to the deficiencies of the union bound (see Section 2). They therefore provide useful results at rates exceeding the cutoff rate, where union bounds are usually useless. These bounds solely depend on the distance spectra of the considered linear codes. We exemplify the use of these upper bounds for block codes and ensembles of turbo-like codes. Along their presentation, the underlying connections which exist between these bounds are also demonstrated.

The dependence of the various upper bounds considered in this section on the distance spectra and input-output weight enumerators of the codes is a pleasing property; this makes the bounds attractive for their application to the performance evaluation of various linear codes and ensembles. The reader is referred to the literature which considers various methods for the calculation of the distance spectra and input-output weight enumerators of codes and ensembles (see [14, 38, 45, 46, 66, 155, 190, 216, 220] for the calculation of the distance spectra of algebraic block codes, [37, 75, 107, 112, 127, 146, 147, 168, 174, 207, 222] for convolutional codes, [43, 162] for trellis codes, [19, 18, 17, 28, 56, 92, 97, 142, 151, 159, 161, 160, 163, 171, 174, 191, 208, 217] for turbo codes, [21, 34, 47, 48, 49, 68, 81, 95, 121, 202] for regular and irregular LDPC codes, [36, 65, 196, 195] for product codes, and [1, 54, 87, 93, 117, 118, 123, 150, 148, 166, 211] for accumulate-based codes, and [51, 72] for protograph LDPC codes).

4

Gallager-Type Upper Bounds: Variations, Connections and Applications

Overview: In addressing the Gallager bounds and their variations, we focus on the Duman and Salehi variation which originates from the standard Gallager bound. A large class of efficient recent bounds (or their Chernoff versions) is demonstrated to be a particular instance of the generalized second version of the Duman and Salehi bounds. Implications and applications of these observations are pointed out, including the fully interleaved fading channel, resorting to either matched or mismatched decoding. The proposed approach can be generalized to geometrically uniform non-binary codes, finite state channels, bit interleaved coded modulation systems, and to upper bounds on the conditional decoding error probability.

4.1 Introduction

The Fano [71] and Gallager [82] upper bounds were introduced as efficient tools to determine the error exponents of the ensemble of random codes, providing informative results up to the ultimate capacity limit. Since the advent of information theory, the search for efficient coding systems has motivated the introduction of efficient bounding techniques tailored to specific codes or some carefully chosen ensembles of codes. A classical example is the adaptation of the Fano upper bounding technique [71] to specific codes, as reported in the seminal dissertation by Gallager [81] (to be referred to as the 1961 Gallager-Fano bound).

In continuation to the previous section, we consider here various reported upper bounds on the ML decoding error probability and demonstrate the underlying connections that exist between them; the bounds are based on the distance spectra or the input-output weight enumerators of the considered codes. In the following, we consider variations on the Gallager bounds, and, in particular, we focus on the second version of the recently introduced bounds by Duman and Salehi ([62], [60]) whose derivation is based on the 1965 Gallager bounding technique ([82], [83]). Though originally derived for binary signaling over an additive white Gaussian noise (AWGN) channel, we demonstrate here its considerable generality and show that it provides the natural bridge between the 1961 and 1965 Gallager bounds ([81], [82]). It is suitable for both random and specific codes [50], as well as for either bit or block error probability analysis. It is also demonstrated here that a large class of efficient recent bounds or their Chernoff versions are special cases of the generalized second version of the Duman and Salehi (DS2) bound. We exemplify the use of this generalized bound in various settings, such as the fully interleaved fading channel ([172],[173]). In an important contribution, Divsalar [50] has introduced some efficient and easily applicable bounds, and has also provided insightful observations on the Duman and Salehi bounding technique ([62], [60]) in view of other bounds. In our setting, we shall rely on some of the interesting observations in [50].

The section is organized as follows: The 1965 Gallager bound ([82], [83]) and the DS2 bound [60] are presented in Section 4.2. These two upper bounds form the underlying bounding technique in this section, as we rely on them throughout. In Section 4.3, the 1961 Gallager-Fano bound [81] is presented, and some interconnections among the Gallager bounds and the DS2 bound are demonstrated. It is shown in Section 4.3 that the DS2 bound provides the natural bridge between the 1961 and 1965 Gallager bounds. In Section 4.4, it is demonstrated that many reported bounds on the ML decoding error probability (which were originally derived independently) can be considered as special cases of the DS2 bound. The 1965 Gallager random coding bound is extended in Section 4.5 to the mismatched decoding regime, where the decoder operates in a ML fashion, but may use a mismatched metric. These Gallager-type bounds which are derived for the mismatched decoding regime can be applied to deterministic codes and ensembles. Some reported results ([85], [106], [132]) are derived in Section 4.5, based on an alternative approach which appropriately limits the code ensemble. Some applications and examples of these bounds are presented in Section 4.7, which include fully interleaved fading channels and mismatched metrics. Finally, Section 4.8 concludes the section.

4.2 Gallager bounds for symmetric memoryless channels

4.2.1 The 1965 Gallager bound

Suppose an arbitrary codeword $\underline{x}^m$ (of length-N) is transmitted over a channel. Let $\underline{y}$ designate the observation vector (of N components), and $p_N(\underline{y}|\underline{x}^m)$ be the channel transition probability measure. Then, the conditional ML decoding error probability is given by

$$P_{\mathrm{e}|m} = \sum_{\underline{y}\colon \left\{\exists\, m' \neq m\colon\, p_N(\underline{y}|\underline{x}^{m'}) \geq p_N(\underline{y}|\underline{x}^m)\right\}} p_N(\underline{y}|\underline{x}^m).$$

If the observation vector $\underline{y}$ is such that there exists $m' \neq m$ so that $p_N(\underline{y}|\underline{x}^{m'}) \geq p_N(\underline{y}|\underline{x}^m)$, then for arbitrary $\lambda, \rho \geq 0$, the value of the expression

$$\left(\sum_{m' \neq m} \left(\frac{p_N(\underline{y}|\underline{x}^{m'})}{p_N(\underline{y}|\underline{x}^m)} \right)^{\lambda} \right)^{\rho} \tag{4.1}$$

is clearly lower bounded by 1, and in general, it is always non-negative. The 1965 Gallager bound [82, 83] therefore states that

$$P_{\mathrm{e}|m} \leq \sum_{\underline{y}} p_N(\underline{y}|\underline{x}^m) \left(\sum_{m' \neq m} \left(\frac{p_N(\underline{y}|\underline{x}^{m'})}{p_N(\underline{y}|\underline{x}^m)} \right)^{\lambda} \right)^{\rho}, \quad \lambda, \rho \geq 0\,. \tag{4.2}$$

The upper bound (4.2) is usually not easily evaluated in terms of basic features of particular codes, except for example, orthogonal codes and the special case of $\rho = 1$ and $\lambda = \frac{1}{2}$ (which yields the Bhattacharyya-union bound).

For discrete, memoryless and output-symmetric channels, the upper bound (4.2) is not directly applicable for actual code performance calculation, since it cannot be factored into single-letter expressions (because of the ρth power which operates on the *inner* summation of (4.2)). Therefore, the bound does not lend itself to code performance calculation in terms of the distance spectrum of the ensemble of codes. This difficulty could be circumvented had the ρth power in (4.2) been taken over the expression $\sum_{\underline{y}} \left\{ p_N(\underline{y}|\underline{x}^m) \sum_{m' \neq m} \left(\frac{p_N(\underline{y}|\underline{x}^{m'})}{p_N(\underline{y}|\underline{x}^m)} \right)^{\lambda} \right\}$ (instead of its actual location in the inner summation of the bound (4.2)).

For ensembles of random block codes, where the codewords are independently selected with an arbitrary probability distribution $q_N(\underline{x})$, Gallager derived an upper bound on the average ML decoding error probability (where the average is over the randomly and independently selected codewords). By setting $\lambda = \frac{1}{1+\rho}$ in (4.2), we obtain that for $\rho \geq 0$, the average ML decoding error probability satisfies

$$P_{\mathrm{e}|m} \leq \sum_{\underline{x}^1} \dots \sum_{\underline{x}^M} \left\{ q_N(\underline{x}^1) \dots q_N(\underline{x}^M) \right.$$
$$\left. \cdot \sum_{\underline{y}} p_N(\underline{y}|\underline{x}^m)^{\frac{1}{1+\rho}} \left(\sum_{m' \neq m} p_N(\underline{y}|\underline{x}^{m'})^{\frac{1}{1+\rho}} \right)^{\rho} \right\}.$$

Without any loss of generality, it is assumed that $m = 1$ (since all the codewords are chosen randomly and independently according to the same input distribution $q_N(\underline{x})$), so by changing the order of summation in the last inequality, we obtain that

$$P_{\mathrm{e}} \leq \sum_{\underline{y}} \sum_{\underline{x}^1} q_N(\underline{x}^1) p_N(\underline{y}|\underline{x}^1)^{\frac{1}{1+\rho}} \sum_{\underline{x}^2} \dots \sum_{\underline{x}^M} q_N(\underline{x}^2) \dots q_N(\underline{x}^M)$$
$$\cdot \left(\sum_{m'=2}^{M} p_N(\underline{y}|\underline{x}^{m'})^{\frac{1}{1+\rho}} \right)^{\rho}, \quad \rho \geq 0.$$

By invoking the Jensen inequality in the last inequality, $E[z^\rho] \le (E[z])^\rho$ for $0 \le \rho \le 1$, we obtain that for $0 \le \rho \le 1$

$$\begin{aligned}
P_{\mathrm{e}} &\le \sum_{\underline{y}} \sum_{\underline{x}^1} q_N(\underline{x}^1)\, p_N(\underline{y}|\underline{x}^1)^{\frac{1}{1+\rho}} \Bigg(\sum_{\underline{x}^2} \dots \sum_{\underline{x}^M} q_N(\underline{x}^2) \dots q_N(\underline{x}^M) \\
&\qquad\qquad\qquad\qquad\qquad \cdot \sum_{m'=2}^{M} p_N(\underline{y}|\underline{x}^{m'})^{\frac{1}{1+\rho}} \Bigg)^{\rho} \\
&= \sum_{\underline{y}} \sum_{\underline{x}^1} q_N(\underline{x}^1)\, p_N(\underline{y}|\underline{x}^1)^{\frac{1}{1+\rho}} \left(\sum_{m'=2}^{M} \sum_{\underline{x}^{m'}} q_N(\underline{x}^{m'})\, p_N(\underline{y}|\underline{x}^{m'})^{\frac{1}{1+\rho}} \right)^{\rho} \\
&= \sum_{\underline{y}} \sum_{\underline{x}^1} q_N(\underline{x}^1)\, p_N(\underline{y}|\underline{x}^1)^{\frac{1}{1+\rho}} \left((M-1) \cdot \sum_{\underline{x}} q_N(\underline{x})\, p_N(\underline{y}|\underline{x})^{\frac{1}{1+\rho}} \right)^{\rho}.
\end{aligned}$$

The 1965 Gallager random coding bound therefore reads

$$P_{\mathrm{e}} \le (M-1)^\rho \sum_{\underline{y}} \left(\sum_{\underline{x}} q_N(\underline{x})\, p_N(\underline{y}|\underline{x})^{\frac{1}{1+\rho}} \right)^{1+\rho}, \quad 0 \le \rho \le 1 \qquad (4.3)$$

where P_{e} designates the average decoding error probability, and M is the number of the codewords. For the particular case of a memoryless channel $\left(p_N(\underline{y}|\underline{x}) = \prod_{i=1}^{N} p(y_i|x_i)\right)$ and a memoryless input distribution $\left(q_N(\underline{x}) = \prod_{i=1}^{N} q(x_i)\right)$, the Gallager random coding bound (4.3) admits the form

$$P_{\mathrm{e}} \le e^{-N \cdot E_{\mathrm{r}}(R, \mathbf{q})} \qquad (4.4)$$

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

$$E_{\mathrm{r}}(R, \mathbf{q}) = \max_{0 \le \rho \le 1} \{E_0(\rho, \mathbf{q}) - \rho R\} \qquad (4.5)$$

where

$$E_0(\rho, \mathbf{q}) \triangleq -\ln \left(\sum_{y} \left(\sum_{x} q(x)\, p(y|x)^{\frac{1}{1+\rho}} \right)^{1+\rho} \right). \qquad (4.6)$$

This error exponent is known as the correct exponential dependence of the decoding error probability for code rates above the critical rate [184, Part I] (see Section 5). In [84], Gallager derived an asymptotic expression for the average decoding error probability, and proved that the random coding error exponent is tight for the average code. The result in [84] shows that the weakness of the random coding bound at rates below the critical rate is due to the fact that the best codes are much better than the average at low rates (and not because of the upper bounding technique for the ensemble average).

From the Gallager random coding bound, ensembles of fully random block codes achieve the channel capacity with an optimal soft-decision ML decoder. Moreover, the randomness of the codewords is not crucial for proving the channel coding theorem, and a requirement of pairwise independence of the codewords is sufficient. The latter condition holds, e.g., for multilevel codes when the encoding at each level is performed independently. Using multiuser information theory, it is proved in [212, Section 3] that ensembles of multilevel codes achieve the channel capacity with an ML decoder (see [212, Section 3]). The random coding error exponent (4.5) serves as an efficient design rule for multilevel codes; it serves for properly selecting the rates of the component codes at each level (see [212, Section 4]).

4.2.2 The DS2 bound

The bounding technique of Duman and Salehi [62, 60] originates from the 1965 Gallager bound. Let $\psi_N^m(\underline{y})$ designate an arbitrary probability measure (which may also depend on the transmitted codeword $\underline{x}^m$). The 1965 Gallager bound (4.2) then yields that

$$\begin{aligned} P_{\mathrm{e}|m} &\leq \sum_{\underline{y}} \psi_N^m(\underline{y})\, \psi_N^m(\underline{y})^{-1}\, p_N(\underline{y}|\underline{x}^m) \left(\sum_{m' \neq m} \left(\frac{p_N(\underline{y}|\underline{x}^{m'})}{p_N(\underline{y}|\underline{x}^m)} \right)^{\lambda} \right)^{\rho} \\ &= \sum_{\underline{y}} \psi_N^m(\underline{y}) \left(\psi_N^m(\underline{y})^{-\frac{1}{\rho}}\, p_N(\underline{y}|\underline{x}^m)^{\frac{1}{\rho}} \sum_{m' \neq m} \left(\frac{p_N(\underline{y}|\underline{x}^{m'})}{p_N(\underline{y}|\underline{x}^m)} \right)^{\lambda} \right)^{\rho} \\ &\qquad\qquad \forall\, \lambda, \rho \geq 0. \end{aligned} \tag{4.7}$$

By invoking the Jensen inequality in (4.7), the DS2 bound results

$$P_{\mathrm{e}|m} \leq \left(\sum_{m' \neq m} \sum_{\underline{y}} p_N(\underline{y}|\underline{x}^m)^{\frac{1}{\rho}} \, \psi_N^m(\underline{y})^{1-\frac{1}{\rho}} \left(\frac{p_N(\underline{y}|\underline{x}^{m'})}{p_N(\underline{y}|\underline{x}^m)} \right)^{\lambda} \right)^{\rho} , \quad 0 \leq \rho \leq 1, \; \lambda \geq 0. \tag{4.8}$$

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

$$\psi_N^m(\underline{y}) = \frac{G_N^m(\underline{y}) \; p_N(\underline{y}|\underline{x}^m)}{\sum_{\underline{y}} G_N^m(\underline{y}) \; p_N(\underline{y}|\underline{x}^m)} \, . \tag{4.9}$$

The functions $G_N^m(\underline{y})$ and $\psi_N^m(\underline{y})$ are referred to as the unnormalized and normalized tilting measures, respectively. The substitution of (4.9) into (4.8) yields the following upper bound on the conditional ML decoding error probability

$$P_{\mathrm{e}|m} \leq \left(\sum_{\underline{y}} G_N^m(\underline{y}) \; p_N(\underline{y}|\underline{x}^m) \right)^{1-\rho} \cdot \left\{ \sum_{m' \neq m} \sum_{\underline{y}} p_N(\underline{y}|\underline{x}^m) \; G_N^m(\underline{y})^{1-\frac{1}{\rho}} \left(\frac{p_N(\underline{y}|\underline{x}^{m'})}{p_N(\underline{y}|\underline{x}^m)} \right)^{\lambda} \right\}^{\rho} , \quad 0 \leq \rho \leq 1 \, , \; \lambda \geq 0. \tag{4.10}$$

The upper bound (4.10) was also derived in [50, Eq. (62)].

For the case of memoryless channels, and for the choice of $\psi_N^m(\underline{y})$ as $\psi_N^m(\underline{y}) = \prod_{i=1}^{N} \psi^m(y_i)$ (recalling that the function ψ^m may depend on the transmitted codeword $\underline{x}^m$), the upper bound (4.8) is relatively easily evaluated (similarly to the standard union bounds) for linear block codes. In that case, (4.8) is calculable in terms of the distance spectrum of the code, not requiring the fine details of the code structure. Moreover, (4.8) is also amenable to some generalizations, such as for the class of discrete memoryless channels with arbitrary input and output alphabets.

4.2.2.1 A generalization

As a possible generalization to the above Gallager-based bounding technique, let m be the index of the transmitted codeword, and partition the set of the indices $\{m' : m' \neq m\}$ into an arbitrary number J of disjoint subsets $\{\mathcal{S}_m(j)\}_{j=1}^{J}$. Then, as a starting point for utilizing Gallager-based bounds, one may consider replacing the function in (4.1) with the generalized function

$$\sum_{j=1}^{J} \left(\sum_{m' \in \mathcal{S}_m(j)} \left(\frac{p_N(\underline{y}|\underline{x}^{m'})}{p_N(\underline{y}|\underline{x}^{m})} \right)^{\lambda_m(j)} \right)^{\rho_j}, \quad \lambda_m(j), \rho_j \geq 0$$

which as required, is lower bounded by 1 if there is $m \neq m$ so that $p_N(\underline{y}|\underline{x}^{m'}) \geq p_N(\underline{y}|\underline{x}^{m})$, and is otherwise lower bounded by zero. This yields the following generalized version of (4.8):

$$P_{\mathrm{e}|m} \leq \sum_{j=1}^{J} \left[\sum_{m' \in \mathcal{S}_m(j)} \sum_{\underline{y}} \psi_N^m(\underline{y};j)^{1-\frac{1}{\rho_j}} p_N(\underline{y}|\underline{x}^m)^{\frac{1-\lambda_m(j)\rho_j}{\rho_j}} p_N(\underline{y}|\underline{x}^{m'})^{\lambda_m(j)} \right]^{\rho_j}$$

where $\{\psi_N^m(\underline{y};j)\}_{j=1}^{J}$ is an arbitrary set of J probability tilting measures.

Let us consider the particular choice where the disjoint subsets $\{\mathcal{S}_m(j)\}_{j=1}^{J}$ are selected in a way so that each subset includes all the indices m' of the codewords $\underline{x}^{m'}$ whose Hamming distance from the codeword $\underline{x}^{m}$ is constant. In the case where the all-zero codeword is transmitted, this gives a partitioning to constant Hamming weight subcodes; therefore, it reduces in this specific case to the union bound used with respect to the constant Hamming weight subcodes (see (4.45)), and the Gallager-based bound applied for each of these subcodes separately.

4.2.3 The 1961 Gallager-Fano bound

In his monograph on LDPC codes [81], Gallager introduced an upper bound on the ML decoding error probability for block codes operating over an arbitrary MBIOS channel (see [81, Section 3]). This bound depends on the average distance spectrum of the block code (or ensemble of codes), and therefore, it can be applied to ensembles as well as specific codes (as opposed to the 1965 Gallager bound which applies essentially to the ensemble of random block codes). The derivation of this bound is based on Fano's 1961 bounding technique [71] for random codes, which is adapted in [81] to specific codes. The concept of this bound is detailed in the following. Let $\mathcal{C}$ be a block code of length N which is communicated over an MBIOS channel, and let $\underline{x}^m$ be the transmitted codeword (where $m \in \{0, 1, \ldots, M-1\}$). We define the tilted metric

$$D_m(\underline{x}^{m'}, \underline{y}) \triangleq \ln\left(\frac{f_N^m(\underline{y})}{p_N(\underline{y}|\underline{x}^{m'})}\right) \tag{4.11}$$

where $\underline{x}^{m'}$ is an arbitrary codeword of the code $\mathcal{C}$ and $\underline{y}$ is the received vector (of length N) at the output of the channel, $p_N(\underline{y}|\underline{x}) = \prod_{i=1}^N p(y_i|x_i)$ is the conditional transition probability of the MBIOS channel, and $f_N^m(\underline{y})$ is an arbitrary function which is positive if $p_N(\underline{y}|\underline{x}^{m'})$ is positive for any m' (and may also depend on the transmitted message). Note that the metric D_m is in general not computable at the receiver; it is used here as a conceptual tool to evaluate the upper bound on the decoding error probability. If ML decoding is applied, then an error occurs if there exists $m' \neq m$ so that

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

By increasing the value of $D_m(\underline{x}^m, \underline{y})$, then if the above condition holds, it is likely to hold for more than a single value of m' (where $m' \neq m$). The union bound which upper bounds the probability of the union of events by the sum of the probabilities of the individual events does not therefore yield a tight upper bound, since it counts one decoding error many times in the bound. Gallager therefore chose to separate the set of the observation vectors (of length N), $\mathcal{Y}^N$, into two disjoint subsets: the good subset ($\mathcal{Y}_g^N$) corresponds to the case where the value of $D_m(\underline{x}^m, \underline{y})$ is not larger than a certain threshold which was chosen to be linearly proportional to the block length N (say Nd, where $d \in \mathbb{R}$ is arbitrary), and the bad subset ($\mathcal{Y}_b^N$) corresponds to the case where the value of $D_m(\underline{x}^m, \underline{y})$ is larger than Nd. In the derivation of his bound, Gallager applied the union bound only to the case where the received vector is inside the good region (i.e., if $\underline{y} \in \mathcal{Y}_g^N$). Therefore, we obtain that

$$\begin{aligned}
&\mathcal{Y}^N = \mathcal{Y}_g^N \bigcup \mathcal{Y}_b^N \\
&\mathcal{Y}_g^N \triangleq \Big\{\underline{y} \in \mathcal{Y}^N : D_m(\underline{x}^m, \underline{y}) \leq Nd\Big\} \qquad (4.12)\\
&\mathcal{Y}_b^N \triangleq \Big\{\underline{y} \in \mathcal{Y}^N : D_m(\underline{x}^m, \underline{y}) > Nd\Big\}
\end{aligned}$$

where d is an arbitrary real number. The parameter d is later optimized, so as to get the tightest upper bound within this family.

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

$$P_{\mathrm{e}|m} = \Pr\big(\text{error }, \underline{y} \in \mathcal{Y}_b^N\big) + \Pr\big(\text{error }, \underline{y} \in \mathcal{Y}_g^N\big) \tag{4.13}$$

which leads to the following upper bound:

$$P_{\mathrm{e}|m} \leq \Pr\big(\underline{y} \in \mathcal{Y}_b^N\big) + \Pr\big(\text{error }, \underline{y} \in \mathcal{Y}_g^N\big) \tag{4.14}$$

which resembles the methodology as introduced by Fano [71].

The arbitrary function f_N^m in (4.11) has clearly no effect on the RHS of (4.13) because it does not affect which codeword is decoded when $\underline{x}^m$ is transmitted. However, the function f_N^m affects the upper bound on the conditional ML decoding error probability in (4.14). As was explained in Section 3.1, an inequality like (4.14) forms the starting point of many efficient bounds, e.g., the tangential bound of Berlekamp [22] (see Section 3.2.3 here), Hughes' bound [98] (see Section 3.2.8 on p. 48), the TSB of Poltyrev [152] (see Section 3.2.1 on p. 23), and the bound of Engdahl and Zigangirov [67] (see Section 3.2.6 in p. 44). In the Gallager-Fano approach, the regions $\mathcal{Y}_g^N$, $\mathcal{Y}_b^N$ are related to the choice of the arbitrary function f_N^m in (4.11). At this stage, Fano [71] proceeded with the Chernoff bounding technique and the random coding approach, while the Gallager bound [81] is better suited to treat particular codes.

Based on the Chernoff bound

$$\Pr(\underline{y} \in \mathcal{Y}_b^N) \leq E(e^{sW}) \,, \quad s \geq 0 \tag{4.15}$$

where

$$W \triangleq D_m(\underline{x}^m, \underline{y}) - Nd = \ln\left(\frac{f_N^m(\underline{y})}{p_N(\underline{y}|\underline{x}^m)}\right) - Nd. \tag{4.16}$$

The second term in (4.14) is also upper bounded by a combination of the union and the Chernoff bounds, which then yields

$$\begin{aligned}
&\Pr(\text{decoding error, } \underline{y} \in \mathcal{Y}_g^N) \\
&\quad = \Pr\Big(D_m(\underline{x}^{m'}, \underline{y}) \leq D_m(\underline{x}^m, \underline{y}) \text{ for some } m' \neq m \,, \quad \underline{y} \in \mathcal{Y}_g^N\Big) \\
&\quad \leq \sum_{m' \neq m} \Pr\Big(D_m(\underline{x}^{m'}, \underline{y}) \leq D_m(\underline{x}^m, \underline{y}) \,, \, D_m(\underline{x}^m, \underline{y}) \leq Nd\Big) \\
&\quad \leq \sum_{m' \neq m} E\Big(\exp(t\, Z_{m'} + r\, W)\Big), \quad \forall\, t, r \leq 0
\end{aligned} \tag{4.17}$$

where, based on (4.11)

$$Z_{m'} \triangleq D_m(\underline{x}^{m'}, \underline{y}) - D_m(\underline{x}^m, \underline{y}) = \ln\left(\frac{p_N(\underline{y}|\underline{x}^m)}{p_N(\underline{y}|\underline{x}^{m'})}\right)$$

and W is defined in (4.16). In [81], Gallager considered the case where the functions $f_N^m(\underline{y})$ can be expressed in the product form

$$f_N^m(\underline{y}) = \prod_{i=1}^{N} f(y_i) \tag{4.18}$$

(where the function f does not depend here on the index i). For simplifying the derivation of the 1961 Gallager-Fano bound [81], it was also assumed that the non-negative function f is *even*, i.e., $f(y) = f(-y)$ for all $y \in \mathcal{Y}$. Let $p_x(y) \triangleq p(y|x)$ be the probability transition measure of the MBIOS channel (so $p_1(y) = p_0(-y)$ for all $y \in \mathcal{Y}$).

Let $N_m(l)$ be the number of codewords in the code $\mathcal{C}$ with Hamming distance l from the transmitted codeword $\underline{x}^m$ (where $l \in \{0, 1, \ldots, N\}$).

For ensembles of block codes, we refer to the expected number of codewords with Hamming distance l from the transmitted codeword, where the expectation is taken over all the codes from this ensemble. We note that for geometrically uniform block codes (e.g., linear block codes or fully random block codes), the value of $N_m(l)$ does not depend on the index of the codeword m. For the general case, let $S_l \triangleq \frac{1}{M} \sum_{m=0}^{M-1} N_m(l)$ be the average distance spectrum of the code (or ensemble) $\mathcal{C}$. Based on (4.14)–(4.17) and by optimally setting $t = \frac{r-1}{2}$, the 1961 Gallager-Fano bound on the average ML decoding error probability (see [81, Eqs. (3.28)–(3.30)]) reads

$$P_{\mathrm{e}} \leq g(s)^N \, \exp(-Nsd) + \sum_{l=0}^{N} \Big\{ S_l \, [h(r)]^l \, [g(r)]^{N-l} \Big\} \exp(-Nrd) \tag{4.19}$$

where $s \geq 0, r \leq 0$ and $d \in \mathbb{R}$. The term in (4.19) for $l = 0$ accounts for the pathological possibility that another codeword in $\mathcal{C}$ is identical to the transmitted codeword; S_0 is the (average) number of codewords in $\mathcal{C}$, other than the transmitted codeword, which are identical to it.

From the symmetry of the channel, and since the function f is assumed to be even, Gallager obtained with the Chernoff bounding technique (see (4.15)–(4.17)) that the functions g and h may be expressed as

$$\begin{aligned} g(s) &= \sum_{y \in \mathcal{Y}} p_0(y)^{1-s} \, f(y)^s \\ &= \frac{1}{2} \sum_{y \in \mathcal{Y}} \Big\{ \Big([p_0(y)]^{1-s} + [p_1(y)]^{1-s} \Big) \, f(y)^s \Big\} \\ h(r) &= \sum_{y \in \mathcal{Y}} \Big\{ [p_0(y) \, p_1(y)]^{\frac{1-r}{2}} \, f(y)^r \Big\} \end{aligned} \tag{4.20}$$

and where the sums above are replaced by integrals when the output alphabet is continuous. The upper bound (4.19) on the ML decoding error probability depends on the distance spectrum of the considered block code, and it therefore applies to fixed codes and also structured ensembles of codes. It can be verified (though not mentioned in [81]) that the optimization of the parameter d in (4.19) yields that

$$e^{-Nd} = \left(-\frac{r}{s} \, \frac{1}{[g(s)]^N} \, \sum_{l=0}^{N} S_l [h(r)]^l \, [g(r)]^{N-l} \right)^{\frac{1}{s-r}} \tag{4.21}$$

and by the substitution of (4.21) in the RHS of (4.19), one obtains that the tightest bound within this family gets the expression

$$P_{\mathrm{e}} \le 2^{H(\rho)} \, [g(s)]^{N(1-\rho)} \left(\sum_{l=0}^{N} S_l \, [h(r)]^l \, [g(r)]^{N-l} \right)^{\rho}, \quad 0 \le \rho \le 1. \tag{4.22}$$

Here $\rho \triangleq \frac{s}{s-r}$ (so it follows that $0 \le \rho \le 1$, since $s \ge 0$ and $r \le 0$), and $H(\rho) \triangleq -\rho \log_2(\rho) - (1-\rho)\log_2(1-\rho)$ is the binary entropy function to the base 2. The optimization of the function f in (4.19) yields an implicit solution (as indicated in [81, Eq. (3.40)]). The following approximation to the optimal non-negative and symmetric function f was proposed by Gallager (see [81, Eq. (3.41)]):

$$f(y) = k \left\{ \frac{\left[p_0(y)^{\frac{1-r}{2}} + p_1(y)^{\frac{1-r}{2}} \right]^2}{p_0(y)^{1-s} + p_1(y)^{1-s}} \right\}^{\frac{1}{s-r}} \tag{4.23}$$

where the constant k in (4.23) is arbitrary and cancels out in the bound. For the ensemble of fully random binary block codes where all the binary block codes of a fixed rate and a fixed block length are equiprobable, Gallager noted in his tutorial (see [81, p. 30]) that the function f in (4.23) is indeed the optimal one (so in this case, it is not just an approximation); the optimality of the function f is in the sense of minimizing the 1961 Gallager-Fano upper bound in (4.19) among all the non-negative and even functions (since these properties on the arbitrary function f were required for the derivation of this bound). The optimality of the function f in (4.23) is proved rather easily with calculus of variations, starting from Eqs. (4.19) and (4.20); to this end, one relies on the distance spectrum of the ensemble of fully random binary block codes with rate R bits per channel use and block length N, which is given by

$$S_l = 2^{-N(1-R)} \binom{N}{l}, \quad l = 0, 1, \ldots, N.$$

4.3 Interconnections between bounds

4.3.1 The random coding version of the DS2 bound

We show here that the random coding version of the DS2 bound coincides with the well known 1965 Gallager bound for random codes. For the ensemble of random codes, where the N-length codewords are randomly and independently selected with respect to the input distribution $q_N(\underline{x})$, the DS2 bound yields the following upper bound on the conditional decoding error probability:

$$P_{\mathrm{e}|m} \leq (M-1)^\rho \left(\sum_{\underline{y}} p_N(\underline{y}|\underline{x}^m)\, G_N^m(\underline{y}) \right)^{1-\rho}$$

$$\cdot \left\{ \sum_{\underline{y}} p_N(\underline{y}|\underline{x}^m)\, G_N^m(\underline{y})^{1-\frac{1}{\rho}} \sum_{\underline{x}'} q_N(\underline{x}') \left(\frac{p_N(\underline{y}|\underline{x}')}{p_N(\underline{y}|\underline{x}^m)} \right)^\lambda \right\}^\rho ,$$

$$0 \leq \rho \leq 1,\ \lambda \geq 0. \tag{4.24}$$

The optimal non-negative function $G_N^m(\underline{y})$ minimizing (4.24) is

$$G_N^m(\underline{y}) = \left(\sum_{\underline{x}'} q_N(\underline{x}') \left(\frac{p_N(\underline{y}|\underline{x}')}{p_N(\underline{y}|\underline{x}^m)} \right)^\lambda \right)^\rho . \tag{4.25}$$

The substitution of (4.25) into (4.24) gives

$$P_{\mathrm{e}|m} \leq (M-1)^\rho \sum_{\underline{y}} \left\{ p_N(\underline{y}|\underline{x}^m) \left(\sum_{\underline{x}'} q_N(\underline{x}') \left(\frac{p_N(\underline{y}|\underline{x}')}{p_N(\underline{y}|\underline{x}^m)} \right)^\lambda \right)^\rho \right\} . \tag{4.26}$$

After averaging over the transmitted codeword $\underline{x}^m$, which is randomly selected with respect to the input distribution $q_N(\underline{x}^m)$, (4.26) yields the following upper bound on the ML decoding error probability:

$$P_{\mathrm{e}} \leq (M-1)^\rho \sum_{\underline{y}} \left\{ \left(\sum_{\underline{x}} q_N(\underline{x})\, p_N(\underline{y}|\underline{x})^{1-\lambda\rho} \right) \right.$$

$$\left. \cdot \left(\sum_{\underline{x}'} q_N(\underline{x}')\, p_N(\underline{y}|\underline{x}')^\lambda \right)^\rho \right\} . \tag{4.27}$$

Letting $\lambda = \frac{1}{1+\rho}$, the standard random coding Gallager bound [82] results. It is hence demonstrated that in the standard random coding setting, no penalty is incurred for invoking the Jensen inequality in the optimized DS2 bound (see the bound in (4.8) which is subject to the optimization of the probability tilting measure ψ_N^m).

4.3.2 Relations to the 1961 Gallager-Fano bound

In this subsection we study the relation of the 1961 Gallager-Fano bound [81] to the DS2 bound.

Let us assume that a codeword $\underline{x}^m$ was transmitted. From the Chernoff inequality in Eqs. (4.15) and (4.16) we obtain that for arbitrary $s \geq 0$ and $d \in \mathbb{R}$:

$$\begin{aligned}\Pr\left(\underline{y} \in \mathcal{Y}_b^N\right) &\leq \sum_{\underline{y}} p_N(\underline{y}|\underline{x}^m)\, \exp\Big(s\, \left(D_m(\underline{x}^m, \underline{y}) - Nd\right)\Big) \\ &= e^{-Nsd} \sum_{\underline{y}} p_N(\underline{y}|\underline{x}^m) \left(\frac{f_N^m(\underline{y})}{p_N(\underline{y}|\underline{x}^m)}\right)^s . \qquad (4.28)\end{aligned}$$

Similarly, we obtain from the Chernoff upper bound in (4.17) that given that the codeword $\underline{x}^m$ was transmitted, then the conditional joint probability of having the observation vector in the good region which is associated with the 1961 Gallager-Fano bound (i.e., the event that the received vector $\underline{y}$ is inside $\mathcal{Y}_g^N$), and also that the ML decoder makes an error, satisfies the following inequality for arbitrary $t, r \leq 0$ and $d \in \mathbb{R}$:

$$\begin{aligned}&\Pr\left(\text{decoding error}, \underline{y} \in \mathcal{Y}_g^N\right) \\ &\quad\leq \sum_{m' \neq m} p_N(\underline{y}|\underline{x}^m)\, \exp\left(tZ_{m'} + rW\right) \\ &\quad= e^{-Nrd} \sum_{m' \neq m} \sum_{\underline{y}} p_N(\underline{y}|\underline{x}^m) \left(\frac{p_N(\underline{y}|\underline{x}^m)}{p_N(\underline{y}|\underline{x}^{m'})}\right)^t \left(\frac{f_N^m(\underline{y})}{p_N(\underline{y}|\underline{x}^m)}\right)^r . \qquad (4.29)\end{aligned}$$

Based on the upper bound on the conditional ML decoding error probability (4.14), and from Eqs. (4.28), (4.29), we obtain that

$$P_{\mathrm{e}|m} \le e^{-Nsd} \sum_{\underline{y}} p_N(\underline{y}|\underline{x}^m) \left(\frac{f_N^m(\underline{y})}{p_N(\underline{y}|\underline{x}^m)} \right)^s$$
$$+ e^{-Nrd} \sum_{m' \ne m} \sum_{\underline{y}} p_N(\underline{y}|\underline{x}^m) \left(\frac{p_N(\underline{y}|\underline{x}^m)}{p_N(\underline{y}|\underline{x}^{m'})} \right)^t \left(\frac{f_N^m(\underline{y})}{p_N(\underline{y}|\underline{x}^m)} \right)^r$$

where $s \ge 0$, $r \le 0$, $t \le 0$, and $d \in \mathbb{R}$. In a similar way to the optimization of the parameter d in Section 4.2.3, by doing the same kind of optimization in the bound above, and by substituting $\rho = \frac{s}{s-r}$ $(0 \le \rho \le 1)$ in the resulting optimized bound (with respect to the parameter d), one obtains the following upper bound on the conditional ML decoding error probability:

$$P_{\mathrm{e}|m} \le 2^{H(\rho)} \left(\sum_{\underline{y}} p_N(\underline{y}|\underline{x}^m) \left(\frac{f_N^m(\underline{y})}{p_N(\underline{y}|\underline{x}^m)} \right)^s \right)^{1-\rho}$$
$$\cdot \left(\sum_{m' \ne m} \sum_{\underline{y}} p_N(\underline{y}|\underline{x}^m) \left(\frac{p_N(\underline{y}|\underline{x}^m)}{p_N(\underline{y}|\underline{x}^{m'})} \right)^t \left(\frac{f_N^m(\underline{y})}{p_N(\underline{y}|\underline{x}^m)} \right)^{s\left(1-\frac{1}{\rho}\right)} \right)^\rho,$$
$$0 \le \rho \le 1 \tag{4.30}$$

as was first indicated by Divsalar [50, Eqs. (71), (72)].

Divsalar [50] has renamed $-t$ by λ (where $t \le 0$ is introduced in (4.17), so $\lambda \ge 0$ as required in (4.10)), and has also set

$$G_N^m(\underline{y}) = \left(\frac{f_N^m(\underline{y})}{p_N(\underline{y}|\underline{x}^m)} \right)^s \tag{4.31}$$

deriving then the DS2 bound (4.10) with an additional factor of $2^{H(\rho)}$ (where $1 \le 2^{H(\rho)} \le 2$). This demonstrates the superiority of the DS2 bound over the 1961 Gallager-Fano bounding technique when applied to a particular code or ensemble of codes. It has been demonstrated in [167] that the 1961 Gallager-Fano bound equals the 1965 Gallager random coding bound (4.27) up to the $2^{H(\rho)}$ coefficient, where as shown before, the latter bound agrees with the optimized DS2 bound.

For the derivation of the 1961 Gallager-Fano bound for arbitrary MBIOS channels, the function $f_N^m(\underline{y})$ was assumed to be expressible in the product form of (4.18) (see [81, Section 3]). From the interesting connection between the DS2 bound and the 1961 Gallager-Fano bound, we arrive to the following conclusion with respect to the application of these two bounds to MBIOS channels. Since the un-normalized tilting measure $G_N^m(\underline{y})$ in the DS2 bound (4.10) is an arbitrary non-negative function, and also based on the relation in (4.31), then we conclude that the one-dimensional function f in the RHS of (4.18) should not be in fact an even function (but should be only a non-negative function); this therefore makes the general form of the DS2 bound when applied to MBIOS channels valid under wider conditions, as compared to the 1961 Gallager-Fano bound.

4.3.3 Geometric interpretation of the Gallager-type bounds

The connections between the Gallager-Fano tilting measure and the Duman and Salehi normalized and un-normalized tilting measures (which are designated here by $f_N^m(\underline{y})$, $\psi_N^m(\underline{y})$ and $G_N^m(\underline{y})$, respectively) are indicated in (4.9) and (4.31). We will see that these connections also provide some geometric interpretations of various reported bounds. The measure $f_N^m(\underline{y})$ in the 1961 Gallager-Fano bound, which in general does not imply a product form, entails a geometric interpretation associated with the conditions in the inequalities (4.12), specifying the disjoint regions $\mathcal{Y}_g^N, \mathcal{Y}_b^N \subseteq \mathcal{Y}^N$. The geometric interpretation of the 1961 Gallager-Fano bound is not necessarily unique, as measures $f_N^m(\underline{y})$ of different functional structure may imply equivalent conditions in the inequality

$$\ln\left(\frac{f_N^m(\underline{y})}{p_N(\underline{y}|\underline{x}^m)}\right) \leq Nd. \tag{4.32}$$

The non-uniqueness of the measures $f_N^m(\underline{y})$ in this respect is due to the shifting and factoring invariance of inequality (4.32), and since the parameter d in (4.19) is subjected to optimization. We demonstrate here the connection between the non-unique measure $f_N^m(\underline{y})$ and its associated decision region, and exemplify the non-uniqueness property by focusing on the Divsalar bound [50] (see also Section 3.2.4 on p. 36).

The conditional pdf for the binary-input AWGN channel is

$$p_N(\underline{y}|\underline{x}^m) = \prod_{l=1}^{N} \left\{ \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2} \, (y_l - \gamma x^m(l))^2 \right] \right\} \tag{4.33}$$

where $\gamma \triangleq \sqrt{\frac{2RE_{\rm b}}{N_0}}$. Here $x^m(l)$ designates the l-th symbol of the m-th codeword, where 0 and 1 are mapped to 1 and -1, respectively.

Divsalar bound [50] is specified by the associated decision region

$$\mathcal{Y}_g^N = \left\{ \underline{y} | \sum_{l=1}^{N} \big(y_l - \eta\gamma x^m(l)\big)^2 \leq N r^2 \right\} \tag{4.34}$$

which is an N-dimensional sphere whose center that is located along the line connecting the origin to the codeword $\underline{x}^m$ (see Fig. 3.2 on p. 38). The parameters r, η are analytically optimized in Divsalar bound [50]. It can be verified that the following Gallager-Fano tilting measures in (4.32) imply the same decision region in (4.34):

$$f_N^m(\underline{y}) = \prod_{l=1}^{N} \left\{ \exp\left(\frac{(1-\eta)\, y_l^2}{2\eta} \right) \right\}, \tag{4.35a}$$

$$f_N^m(\underline{y}) = \prod_{l=1}^{N} \left\{ \exp\Big(\gamma\, (1-\eta)\, x^m(l)\, y_l \Big) \right\}, \tag{4.35b}$$

$$f_N^m(\underline{y}) = \prod_{l=1}^{N} \left\{ \exp\Big(\theta \left(y_l - \phi x^m(l)\right)^2 \Big) \right\} \tag{4.35c}$$

$$\text{where } \phi \triangleq \gamma \left(\eta + \frac{\eta - 1}{2\theta} \right), \ \ \theta > -\frac{1}{2}.$$

This can be done by examining the geometrical regions associated with the tilting measures $f_N^m(\underline{y})$ in (4.35)–(4.35c). As an example, we show that the measure $f_N^m(\underline{y})$ in (4.35c) yields the geometrical region in (4.34).

Proof. From Eqs. (4.33) and (4.35c), we get

$$
\begin{aligned}
&\ln\left(\frac{f_N^m(\underline{y})}{p_N(\underline{y}|\underline{x}^m)}\right)\\
&= \ln\left(\frac{\prod_{l=1}^{N}\left\{\exp\left(\theta\left(y_l - \phi x^m(l)\right)^2\right)\right\}}{\prod_{l=1}^{N}\left\{\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(y_l - \gamma x^m(l)\right)^2\right)\right\}}\right)\\
&= \frac{N}{2}\ln(2\pi) + \sum_{l=1}^{N}\left\{\theta\left(y_l - \phi x^m(l)\right)^2 + \frac{1}{2}\left(y_l - \gamma x^m(l)\right)^2\right\}\\
&= \left(\theta + \frac{1}{2}\right)\sum_{l=1}^{N} y_l^2 - (2\theta\phi + \gamma)\sum_{l=1}^{N} x^m(l)\, y_l\\
&\quad + N\left(\theta\phi^2 + \frac{\gamma^2}{2} + \frac{\ln(2\pi)}{2}\right)\\
&= \left(\theta + \frac{1}{2}\right)\sum_{l=1}^{N}\left(y_l - \frac{2\theta\phi + \gamma}{2\theta + 1}\cdot x^m(l)\right)^2\\
&\quad + N\left[\theta\phi^2 + \frac{\gamma^2}{2} + \frac{\ln(2\pi)}{2} - \frac{(2\theta\phi + \gamma)^2}{2(2\theta + 1)}\right].
\end{aligned}
\tag{4.36}
$$

The equivalence between the geometrical regions defined in (4.32) and (4.34) is invariant to a shift in the value of the parameter d in (4.32), and to factoring by a positive constant in (4.34). The reason for this invariance is because the two parameters d and r in (4.32) and (4.34), respectively, are subjected to optimizations. This implies that we need $\theta + \frac{1}{2}$ to be positive, and by comparing (4.34) with (4.36), then we also require the following equation to be satisfied:

$$\frac{2\theta\phi + \gamma}{2\theta + 1} = \gamma\eta$$

which yields the conditions in (4.35c). □

We note that although the three tilting measures $f_N^m(\underline{y})$ in (4.35a)–(4.35c) imply the same decision region $\mathcal{Y}_g^N$, the second and third measures (in (4.35b) and (4.35c), respectively) are amenable to generalizations for the class of fully interleaved fading channels, as opposed to the first measure (since the exponential term of the first tilting measure in (4.35a) is quadratic in $\{y_l\}_{l=1}^N$, but does not depend on the coordinates of the transmitted codeword $\{x^m(l)\}_{l=1}^N$); it therefore demonstrates a different functional behavior of these tilting measures.

The geometrical region which is associated with the sphere bound (see Section 3.2.5 on p. 42) is an N-dimensional sphere whose center coincides with the transmitted codeword (so this region is a particular case of (4.34) where $\eta = 1$). For the sphere bound, the tilting measures in (4.35a) and (4.35b) are identical (i.e., $f_N^m(\underline{y}) \equiv 1$), and the third tilting measure in (4.35c) gets the form

$$f_N^m(\underline{y}) = \prod_{l=1}^{N} \left\{ \exp\Big(\theta \left(y_l - \gamma x^m(l)\right)^2\Big) \right\}, \quad \theta > -\frac{1}{2} \, .$$

4.4 Special cases of the DS2 bound

In this section we demonstrate that *many reported upper bounds can be considered as special cases of the DS2 bound* [1] (see Section 4.2.2 on p. 96). These observations rely on the material presented in [50, 167, 183].

4.4.1 The Shulman and Feder bound

We consider here the transmission of a binary linear block code $\mathcal{C}$ where the communication takes place over a memoryless binary-input output-symmetric (MBIOS) channel. The analysis refers to the decoding error probability under soft-decision ML decoding.

The Shulman and Feder bound (SFB) [187] on the block error probability of an (N, K) binary linear block code $\mathcal{C}$, transmitted over an MBIOS channel is given by

$$P_{\mathrm{e}} \leq 2^{-N E_{\mathrm{r}}\left(R + \frac{\log \alpha(\mathcal{C})}{N}\right)} \tag{4.37}$$

[1] The first version of the Duman and Salehi bounds which is introduced in this section is shown to be a particular case of the DS2 bound. Therefore, we introduced the latter bound in Section 4.2, before the presentation of their first version bound in Section 4.4.

where

$$E_{\mathrm{r}}(R) = \max_{0 \le \rho \le 1} \left(E_0(\rho) - \rho R\right) \tag{4.38}$$

$$E_0(\rho) \triangleq -\log_2 \left\{ \sum_y \left[\frac{1}{2} p(y|0)^{\frac{1}{1+\rho}} + \frac{1}{2} p(y|1)^{\frac{1}{1+\rho}} \right]^{1+\rho} \right\}. \tag{4.39}$$

E_{r} is the random coding error exponent [82], $R \triangleq \frac{K}{N}$ designates the code rate in bits per channel use, and

$$\alpha(\mathcal{C}) \triangleq \max_{1 \le l \le N} \frac{A_l}{2^{-N(1-R)} \binom{N}{l}}. \tag{4.40}$$

In the RHS of (4.40), $\{A_l\}$ denotes the distance spectrum of the code. Hence, for fully random block codes, $\alpha(\mathcal{C})$ is equal to 1, and the SFB particularizes to the random coding bound [82]. In general, the parameter $\alpha(\mathcal{C})$ in the SFB (4.37) measures the maximal ratio of the distance spectrum of a code (or ensemble) and the average distance spectrum which corresponds to fully random block codes of the same block length and rate.

The original proof of the SFB is quite involved. In [183], a simpler proof of the SFB is derived, and by doing so, the simplified proof reproduces the SFB as a particular case of the DS2 bound (see Eq. (4.8)). In light of the significance of the proof concept to the continuation of our paper, we outline this proof briefly.

Since we deal with linear block codes and the communication channel is memoryless, binary-input output-symmetric channel (MBIOS), one can assume without any loss of generality that the all zero codeword $\underline{c}^0$ is the transmitted vector. In order to facilitate the expression of the upper bound (4.10) in terms of distance spectrum of the block code $\mathcal{C}$, we consider here the case where the un-normalized tilting measure $G_N^0(\underline{y})$ can be expressed in the following product form:

$$G_N^0(\underline{y}) = \prod_{i=1}^{N} g(y_i) \tag{4.41}$$

where g is an arbitrary non-negative scalar function, and the channel is by assumption MBIOS, so that the transition probability measure is expanded in the product form

$$p_N(\underline{y}|\underline{c}^{m'}) = \prod_{i=1}^{N} p(y_i|c^{m'}(i)) \tag{4.42}$$

where $\underline{c}^{m'} = (c^{m'}(1), \ldots, c^{m'}(N))$. Hence, the upper bound on the conditional ML decoding error probability given in (4.10) can be rewritten as

$$\begin{aligned}
P_{\mathrm{e}} &= P_{\mathrm{e}|0} \\
&\leq \left(\sum_y g(y)\, p(y|0)\right)^{N(1-\rho)} \left\{\sum_{l=1}^{N} A_l \left(\sum_y g(y)^{1-\frac{1}{\rho}} p(y|0)\right)^{N-l}\right. \\
&\qquad \left. \cdot \left(\sum_y g(y)^{1-\frac{1}{\rho}} p(y|0)^{1-\lambda} p(y|1)^{\lambda}\right)^{l}\right\}^{\rho} \qquad \begin{array}{c}\lambda \geq 0, \\ 0 \leq \rho \leq 1\end{array} \\
&\leq \left(\max_{0<l\leq N} \frac{A_l}{2^{-N(1-R)} \binom{N}{l}}\right)^{\rho} \left(\sum_y g(y)\, p(y|0)\right)^{N(1-\rho)} 2^{-N(1-R)\rho} \\
&\qquad \cdot \left\{\sum_y g(y)^{1-\frac{1}{\rho}} p(y|0) + \sum_y g(y)^{1-\frac{1}{\rho}} p(y|0)^{1-\lambda} p(y|1)^{\lambda}\right\}^{N\rho}.
\end{aligned} \tag{4.43}$$

By setting

$$g(y) = \left[\frac{1}{2} p(y|0)^{\frac{1}{1+\rho}} + \frac{1}{2} p(y|1)^{\frac{1}{1+\rho}}\right]^{\rho} p(y|0)^{-\frac{\rho}{1+\rho}}, \quad \lambda = \frac{1}{1+\rho} \tag{4.44}$$

and using the symmetry of the channel (where $p(y|0) = p(-y|1)$), the SFB follows readily.

Miller and Burshtein suggested in [133, Theorem 1] to combine the union bound and the SFB in a sophisticated manner. The idea of combining the union bound and the SFB is based on using the union bound at the portion of the Hamming weights where the distance spectrum of the considered ensemble deviates considerably from the one which corresponds to the ensemble of fully random block codes, and using the SFB for the complementary portion of the Hamming weights where the two distance spectra are close enough.

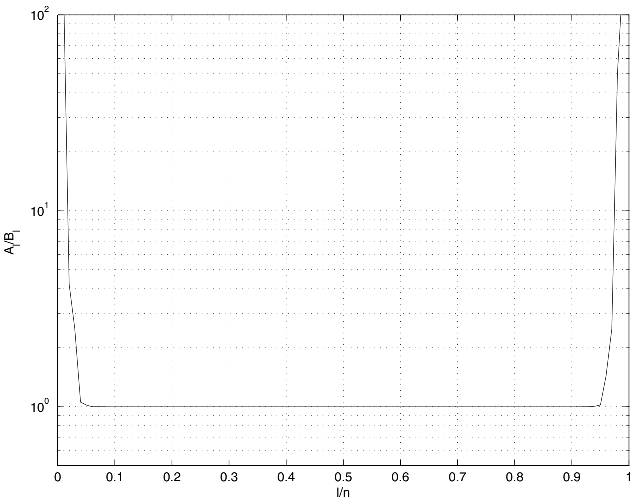

Fig. 4.1 The ratio between the average distance spectrum of a random ensemble of turbo-block codes $\{A_l\}$ and the binomial distribution $\{B_l\}$ which characterizes the average distance spectrum of fully random block codes with the same block length and code rate. We consider an ensemble of uniformly interleaved turbo-block codes where the two constituent codes are binary, linear and systematic block codes of the same rate and block length which are chosen at random. The overall rate of the ensemble is $R = 0.72$ bits per channel use, and the block length of the turbo code is $N = 100$ bits (this implies that the constituent systematic codes are of dimension 72, and their number of parity bits is 14).

Fig. 4.1 shows the ratio between the average distance spectrum of a random ensemble of turbo-block codes $\{A_l\}$ and the binomial distribution $\{B_l\}$ which characterizes the average distance spectrum of fully random block codes with the same block length and code rate. According to this figure, the union bound is used with respect to the small and large Hamming weights (where the distance spectrum of the ensemble is considerably larger than the corresponding binomial distribution), and the SFB is used with respect to the rest of the Hamming weights (i.e., the intermediate values for which the two distance spectra are close enough, so that the ratio depicted in this figure is close to 1).

An upper bound on the ML decoding error probability which combines the SFB with the union bound was used for the analysis of LDPC codes under ML decoding (see [133] and [176, Theorem 2.2]). This generalized bound was used in order to prove that for an arbitrary

MBIOS channel, properly chosen ensembles of regular LDPC codes achieve under ML decoding rates which can be made arbitrarily close to the channel capacity.[2] Based on the same bounding technique which combines the union bound with the SFB, the performance of punctured LDPC codes under ML decoding was studied in [95]. It is shown in this paper that under ML decoding, capacity-achieving codes of any rate and for any MBIOS channel can be constructed by puncturing some original LDPC codes with small enough rate; this indicates the high potential of rate-compatible puncturing to be used in designing capacity-achieving codes for an arbitrary MBIOS channel (see also [88] for the design and analysis of punctured LDPC codes under iterative message-passing decoding). These results demonstrate that the generalized bound which properly combines the SFB with the union bound provides a rather powerful bounding technique for the ML analysis of various ensembles of LDPC codes (or other capacity-approaching ensembles of linear codes). We refer the reader to [21] where Burshtein and Bennatan stated this bounding technique for ensembles of non-binary codes, and applied their bounds to various ensembles of non-binary LDPC codes which are communicated over an arbitrary discrete memoryless channel.

The SFB was recently improved by Twitto et al. [201, 200] where the general approach of this improvement relies on variations of the DS2 bound which tighten the pre-exponent of the SFB. The improved bounding technique was adapted for the analysis of the block error probability as well as the bit error probability of binary linear block codes. For codes of high rate, the improved version of the SFB outperforms the tangential-sphere bound, as exemplified in [201, 200] for some turbo-like ensembles.

[2] We note that a parallel result was not proved yet for any ensemble of LDPC codes under a sub-optimal iterative message-passing decoding algorithm. For ensembles of regular LDPC codes, it is well known that by increasing the degrees of the variable nodes and the parity-check nodes so that the design rate stays constant, the achievable rates of ensembles of regular LDPC codes improve under ML decoding, though their achievable rates under iterative message-passing decoding degrade.

4.4.2 Gallager-type bounds for the binary-input AWGN channel

In this section, we present various upper bounds on the ML decoding error probability for the binary-input AWGN channels; the derivation of these bounds rely on the DS2 bound which is a variation of the Gallager bounding technique. We show that these reported bounds are special cases of the DS2 bound. In general, we prove it by choosing appropriate probability tilting measures which enable to derive these bounds as special cases of the DS2 bound.

Let $\mathcal{C}$ be a binary linear block code of length N. Throughout this section, $\delta \triangleq \frac{d}{N}$ (where $0 \leq \delta \leq 1$) designates the normalized Hamming weight of an arbitrary codeword of $\mathcal{C}$ whose Hamming weight is equal to d, and $r_N(\delta) \triangleq \frac{\ln S_d}{N}$ designates the exponential growth rate of the distance spectrum of $\mathcal{C}$ as a function of δ.

The derivation of the upper bounds in this section relies on the partitioning of the code $\mathcal{C}$ into constant Hamming weight subcodes $\{\mathcal{C}_d\}_{d=0}^N$ where the subcode $\mathcal{C}_d$ includes all the codewords of Hamming weight d, and also the all-zero codeword. We note that although the subcodes $\{\mathcal{C}_d\}$ are not necessarily linear (since the sum of two codewords of Hamming weight d is not necessarily a codeword of such a Hamming weight), the upper bound on the *conditional* error probability (4.8) (see p. 97) of every subcode $\mathcal{C}_d$ is still valid; the reason is that the linearity of the code is not required for the validity of (4.8).

Due to the channel symmetry and the linearity of the code, an overall union bound over the subcodes yields that

$$P_{\mathrm{e}} = P_{\mathrm{e}|0} \leq \sum_{d=d_{\min}}^{N} P_{\mathrm{e}|0}(d) \tag{4.45}$$

where $d_{\min}$ is the minimum Hamming distance of the code $\mathcal{C}$, and $P_{\mathrm{e}|0}(d)$ is the conditional ML decoding error probability of the subcode $\mathcal{C}_d$, given that the all-zero codeword is transmitted. In the continuation to this section, we will apply the DS2 bound with different tilting measures in order to re-derive previously reported upper bounds as special cases of the former bound. From (4.45), it suffices to obtain bounds on $P_{\mathrm{e}|0}(d)$.

(1) 4.4.2.1 Duman and Salehi bound (first version)

The Duman and Salehi bound in [62] (which is named here as the first version of the Duman and Salehi bounds) is a special case of the DS2 bound [60]. More specifically, the bound in [62] is a special case of the generalized DS2 bound in (4.8) where the normalized tilting measure is given by

$$\psi_N^m(\underline{y}) = \prod_{l=1}^{N} \left\{ \sqrt{\frac{\alpha}{2\pi}} \exp\left[-\frac{\alpha}{2} \left(y_l - \frac{\beta}{\alpha} \sqrt{\frac{2E_{\text{s}}}{N_0}}\, x^m(l) \right)^2 \right] \right\}$$
$$\alpha > 0, \ \ \beta \in \mathbb{R}. \qquad (4.46)$$

We assume here that the components of the codeword $\underline{x}^m$ are either $+1$ or -1 for a '0' or '1' input, respectively. By (4.8), we obtain the following upper bound on $P_{\text{e}|0}(d)$:

$$P_{\text{e}|0}(d) \le (S_d)^\rho \left(\int_{-\infty}^{+\infty} p_0(y)^{\frac{1}{\rho}}\, \psi(y)^{1-\frac{1}{\rho}}\, dy \right)^{(N-d)\rho}$$
$$\cdot \left(\int_{-\infty}^{+\infty} p_0(y)^{\frac{1-\lambda\rho}{\rho}}\, p_1(y)^{\lambda}\, \psi(y)^{1-\frac{1}{\rho}}\, dy \right)^{d\rho}$$
$$0 \le \rho \le 1, \ \lambda > 0 \qquad (4.47)$$

where from (4.46), and given that the all-zero codeword is transmitted, then we obtain that

$$\psi(y) = \sqrt{\frac{\alpha}{2\pi}} \exp\left[-\frac{\alpha}{2} \left(y - \frac{\beta}{\alpha} \sqrt{\frac{2E_{\text{s}}}{N_0}} \right)^2 \right], \quad y \in \mathbb{R} \quad (4.48)$$

and the probability density functions

$$p_0(y) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2} \left(y - \sqrt{\frac{2RE_{\text{b}}}{N_0}} \right)^2 \right],$$
$$p_1(y) = p_0(-y), \quad y \in \mathbb{R} \qquad (4.49)$$

refer to the binary-input AWGN channel. The substitution of Eqs. (4.48) and (4.49) into (4.47), and the optimization over the parameters $\lambda \geq 0$ and $\beta \in \mathbb{R}$ in this bound gives (see [62])

$$\beta^* = \frac{1 - \frac{d}{N}}{\frac{1}{\alpha} - \frac{d}{N}(1-\rho)}, \quad \lambda^* = \frac{1}{2}\left(\beta^* + \frac{1-\beta^*}{\rho}\right). \tag{4.50}$$

The closed-form expressions of the integrals in (4.47) for the specific probability measures ψ and p_0 in (4.48) and (4.49), respectively, gives the following upper bound:

$$\begin{aligned} P_{\mathrm{e}|0}(d) \leq (S_d)^\rho \, \alpha^{\frac{N(1-\rho)}{2}} & \left(\alpha - \frac{\alpha-1}{\rho}\right)^{-\frac{N\rho}{2}} \\ & \exp\left\{\frac{NRE_{\mathrm{b}}}{N_0}\left[-1 + \frac{(\beta^*)^2(1-\rho)}{\alpha}\right.\right. \\ & \left.\left. + \frac{\rho\left(1-\frac{d}{N}\right)\left(\beta^* + \frac{1-\beta^*}{\rho}\right)^2}{\alpha - \frac{\alpha-1}{\rho}}\right]\right\}. \end{aligned} \tag{4.51}$$

The bound in (4.51) is the first version of the Duman and Salehi bounds [62], and it is numerically optimized over the range $0 < \alpha < \frac{1}{1-\rho}$, $0 < \rho \leq 1$ (we note that the reason for the range of α as above is because we get from (4.51) that $\alpha - \frac{\alpha-1}{\rho}$ should be positive, and also $\alpha > 0$ is a requirement in (4.46)).

Based on our notation for the normalized Hamming weight (δ) and the exponential growth rate of the distance spectrum ($r_N(\delta)$) of a linear block code, then the bound in (4.47) is also expressible in the following equivalent form:

$$\begin{aligned} P_{\mathrm{e}|0}(d) \leq \Bigg\{ & e^{\rho r_N(\delta)} \left(\int_{-\infty}^{+\infty} p_0(y)^{\frac{1}{\rho}} \, \psi(y)^{1-\frac{1}{\rho}} \, dy\right)^{(1-\delta)\rho} \\ & \cdot \left(\int_{-\infty}^{+\infty} p_0(y)^{\frac{1-\lambda\rho}{\rho}} \, p_1(y)^\lambda \, \psi(y)^{1-\frac{1}{\rho}} \, dy\right)^{\delta\rho} \Bigg\}^N \\ & 0 \leq \rho \leq 1, \; \lambda > 0. \end{aligned} \tag{4.52}$$

This form is useful for the discussion on the $\frac{E_{\mathrm{b}}}{N_0}$ – thresholds of linear codes (see Section 4.7.3 on p. 141).

(2) 4.4.2.2 Viterbi and Viterbi bound (first version)

The Viterbi and Viterbi bound [209] is an upper bound on the ML decoding error probability for BPSK modulated block codes operating over a binary-input AWGN channel. It can be verified that it is also a particular case of the first version of Duman and Salehi bounds by substituting $\alpha = 1$ in (4.50) and (4.51), which yields the following conditional upper bound on the ML decoding error probability:

$$P_{\mathrm{e}|0}(d) \leq (S_d)^{\rho} \exp\left(-\frac{NRE_{\mathrm{b}}}{N_0} \frac{\left(\frac{d}{N}\right)\rho}{1-\frac{d}{N}(1-\rho)}\right), \quad 0 \leq \rho \leq 1\,.$$

The optimization over the parameter ρ then gives the Viterbi & Viterbi upper bound [209], which reads

$$P_{\mathrm{e}|0}(d) \leq \exp\Big(-N\,E_{v_1}(\delta)\Big)$$

where

$$E_{v_1}(\delta) = \begin{cases} \delta c - r_N(\delta), & 0 \leq \frac{r_N(\delta)}{c} \leq \delta\,(1-\delta) \\ \left(\sqrt{c} - \sqrt{\frac{(1-\delta)\,r_N(\delta)}{\delta}}\right)^2, & \delta\,(1-\delta) \leq \frac{r_N(\delta)}{c} \leq \frac{\delta}{1-\delta} \end{cases}$$

and

$$\delta \triangleq \frac{d}{N}\,,\;\; r_N(\delta) \triangleq \frac{\ln(S_d)}{N}\,,\;\; c \triangleq \frac{E_{\mathrm{s}}}{N_0} = \frac{RE_{\mathrm{b}}}{N_0}.$$

(3) 4.4.2.3 Viterbi and Viterbi bound (second version)

The second version of the Viterbi and Viterbi bounds [210] is based on the 1961 Gallager-Fano bound, and is valid for an arbitrary MBIOS channel. This bound reads

$$P_{\mathrm{e}|0}(d) \leq \exp\Big(-N\,E_{v_2}(\delta)\Big)$$

where

$$E_{v_2}(\delta) = \max_{0\leq\rho\leq1} \Big\{-\rho\, r_N(\delta) + \delta \ln\Big(\bar{h}(\rho)\Big) + (1-\delta)\ln\Big(\bar{g}(\rho)\Big) \\ -(1-\rho)\ln\Big(\bar{h}(\rho)+\bar{g}(\rho)\Big)\Big\}\ .$$

From (4.20) and (4.23), and the substitution $\rho = \frac{s}{s-r}$ (so since, $s \geq 0, r \leq 0$, then $0 \leq \rho \leq 1$), we obtain that

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

For the binary-input AWGN channel, $p_0(\cdot)$ is introduced in (4.49). Clearly, for channels with continuous output, the sums above should be replaced by integrals.

The second version of the Viterbi & Viterbi bound is again a special case of the DS2 bound, as noticed by the substitution in (4.10) of the un-normalized tilting measure

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

(4) 4.4.2.4 Divsalar bound

The geometric interpretation of the bound of Divsalar [50] is shown in Fig. 3.2 (see p. 38). We show in Section 4.3.3 (see p. 107) that the spherical region which is shown in Fig. 3.2 is equivalently represented by the condition in (4.32) with an appropriate (and non-unique) tilting measure of the 1961 Gallager-Fano bound (for some possible choices of this tilting measure, see e.g., (4.35a)–(4.35c)).

The following connection between the Gallager-Fano tilting measure ($f_N^m(\underline{y})$) and the normalized tilting measure ($\psi_N^m(\underline{y})$) in the DS2 bound follows from (4.9) and (4.31):

$$\psi_N^m(\underline{y}) = \frac{[f_N^m(\underline{y})]^s\,[p_N(\underline{y}|\underline{x}^m)]^{1-s}}{\sum_{\underline{y}} \left\{ [f_N^m(\underline{y})]^s\,[p_N(\underline{y}|\underline{x}^m)]^{1-s} \right\}} . \tag{4.53}$$

Substituting the tilting measure (4.35c) and the *pdf* (4.33) into (4.53), yields that the corresponding normalized tilting measure (ψ_N^m) for the Divsalar bound coincides with the one in (4.46). This result is verified by setting the parameters as follows:

$$\alpha = 1 - (2\theta + 1)s\,, \quad \beta = 1 - (2\theta + 1)\,\eta s. \tag{4.54}$$

This is the grounds for the observation in [50] that Divsalar bound is a closed form expression of the Duman and Salehi (first version) bound [62].

(5) 4.4.2.5 The Engdahl and Zigangirov bound

For the Engdahl and Zigangirov bound [67] (see Section 3.2.6 on p. 44) which was derived for a binary-input AWGN channel, the decision region $\mathcal{Y}_g^N$ associated with the transmitted codeword $\underline{x}^m = \bigl(x^m(1), x^m(2), \ldots, x^m(N)\bigr)$ is an N-dimensional region whose boundary is a plane

$$\mathcal{Y}_g^N = \left\{ \underline{y} \,|\, \sum_{l=1}^{N} y_l\, x^m(l) \geq N d \right\} \tag{4.55}$$

where $d \in \mathbb{R}$ is a parameter of the 1961 Gallager-Fano bound (with a slight abuse of notation, we make it clear in our discussion when the parameter d stands for the real parameter in the 1961 Gallager-Fano bound (see, e.g., (4.19) and (4.32)), and when d designates the Hamming weights of codewords of a linear code).

The motivation for (4.55) is that inside the good region (i.e., $\underline{y} \in \mathcal{Y}_g^N$), the correlation between the received vector $\underline{y}$ and the transmitted codeword $\underline{x}^m$ is expected to be significant. Therefore, it is above a certain threshold (Nd), where the parameter d is to be optimized, so as to get the tightest upper bound within the family of bounds (4.32) and the associated decision region $\mathcal{Y}_g^N$ in (4.55). The following Gallager-Fano tilting measure can be associated with the same decision region $\mathcal{Y}_g^N$:

$$f_N^m(\underline{y}) = \prod_{l=1}^{N} \left\{ \exp\left(-\frac{1}{2}\, y_l^2 + \beta\, y_l\, x^m(l) \right) \right\} \tag{4.56}$$

where $\beta \neq \gamma$ (γ is introduced in (4.33)). The free parameter β in (4.56) demonstrates again that there might be some functionally different Gallager-Fano tilting measures which imply the same decision region (based on the definition of the regions in (4.12)). It is interesting to note, that for this specific choice of a decision region (4.55), it was demonstrated in [67] that there is no need to invoke the Chernoff bounds for the binary-input AWGN channel, and the two terms in the right hand side of (4.14) can be exactly calculated.

(6) 4.4.2.6 The Chernoff version of various bounds are Gallager-type bounds

In his paper [50], Divsalar derived simplified Chernoff versions of some upper bounds, which are obtained as special instances of the 1961 Gallager-Fano bounding technique. These simplified Chernoff versions include the following bounds:
The Chernoff version of the TSB (see Section 3.2.10.1 on p. 59): The TSB of Poltyrev [152] is one of the tightest known upper bound for block codes which are transmitted over a binary-input AWGN channel and ML decoded (see, e.g., [170], [169]). However, in the random coding setting, it fails to reproduce the random coding exponent [152] while the DS2 bound does. This bound involves a numerical solution of an associated optimization equation ([170, Eq. (4.8)]), and it is therefore not expressed in closed form. In his paper [50], Divsalar derived a simplified Chernoff version of the TSB. That upper bound is shown in [50] to have the same error exponent as the TSB of Poltyrev, and therefore the loosening of the TSB by invoking the Chernoff bounding technique,

does not carry any implications on the tightness of the bound for asymptotically infinite block length.

The Chernoff version of the tangential bound (see Section 3.2.10.2 on p. 66): This version coincides with the first version of the Viterbi and Viterbi bound [209]. It can be shown that the relevant Gallager-Fano decision region is a plane, which is also the case for the Engdahl and Zigangirov bound [67]. However, as noted above, in the latter bound the Chernoff bounding technique is not invoked, improving thus the bound for a finite block length.

The Chernoff version of the sphere bound (see Section 3.2.10.3 on p. 68): This bound results as a particular case of the decision region in Divsalar bound (4.35), where $\eta = 1$. Due to the connection of the Gallager-Fano tilting measure to the Duman and Salehi variation, it is evident that the Chernoff versions of these bounds can be also viewed as special cases of the DS2 bound. It should be emphasized that the mentioned bounds above were originally developed without resorting to Chernoff bounding technique, yielding thus tighter versions of these bounds.

4.4.3 Gallager-type bounds for fully interleaved fading channels with perfect CSI

We demonstrate here various variations of the DS2 bound when applied to the class of fully interleaved Rayleigh fading channels with perfect channel state information at the receiver. This problem is treated in detail in [52], [173], [172] and here we present in a comparative fashion some insightful results. The model is:

$$y = ax + n \tag{4.57}$$

where y stands for the received signal, x stands for the BPSK modulated input signal (that is $\pm\sqrt{2E_{\mathrm{s}}}$) and n designates the additive zero mean and $\frac{N_0}{2}$ variance Gaussian noise component. The fading a is assumed to be perfectly known at the receiver and hence is considered to be real valued, as the receiver compensates for any phase rotation.

Due to the ideal interleaving, the channel is assumed to be memoryless. The bounds are based on first decomposing the code to constant-weight subcodes (where every subcode also includes the all-zero codeword), over which a union bound is invoked as in (4.45).

4.4.3.1 The optimized DS2 bound

Similarly to Section 4.4.2 (see p. 115), the derivation of the upper bounds in this section relies on the partitioning of the code $\mathcal{C}$ into constant Hamming weight subcodes $\{\mathcal{C}_d\}_{d=0}^{N}$ where the subcode $\mathcal{C}_d$ includes all the codewords of Hamming weight d, and also the all-zero codeword. Due to the channel symmetry and the linearity of the code, an overall union bound over the subcodes yields (4.45).

In [172], the measure

$$\Psi(\underline{y}, \underline{a}) = \prod_{l=1}^{N} \psi(y_l, a_l) \tag{4.58}$$

is optimized to yield the tightest conditional DS2 bound (4.8) with respect to the subcode $\mathcal{C}_d$ (given that the all-zero codeword is transmitted), where (y, a) are interpreted as the available measurements at the receiver.

$$\begin{aligned} P_{\mathrm{e}|0}(d) \le (S_d)^{\rho} \Bigg\{ & \left(\int_{-\infty}^{\infty} \int_{0}^{\infty} \psi(y,a)^{1-\frac{1}{\rho}} \, p_0(y,a)^{\frac{1}{\rho}} \, da \, dy \right)^{(1-\delta)\rho} \\ & \cdot \left(\int_{-\infty}^{\infty} \int_{0}^{\infty} \psi(y,a)^{1-\frac{1}{\rho}} p_0(y,a)^{\frac{1-\lambda\rho}{\rho}} p_1(y,a)^{\lambda} \, dady \right)^{\delta\rho} \Bigg\}^{N} \\ & 0 < \rho \le 1, \quad \lambda \ge 0 \end{aligned} \tag{4.59}$$

where $\delta \triangleq \frac{d}{n}$ designates the normalized Hamming weight $(0 \le \delta \le 1)$, the probability density functions p_0 and p_1 are introduced in (3.101) for fully interleaved fading channels, and ψ is an arbitrary non-negative function which satisfies the condition

$$\int_{-\infty}^{\infty} \int_{0}^{\infty} \psi(y,a) \, da \, dy = 1 \ . \tag{4.60}$$

The function $\psi(\cdot,\cdot)$ can be regarded as a tilting measure depending on the measurements y,a, which are perfectly available to the receiver. Resorting to calculus of variations, we obtain the following form for the optimal function ψ, as to provide the tightest upper bound of the family above:

$$\psi(y,a) = \beta\, p_0(y,a) \left(1 + \alpha \left[\frac{p_1(y,a)}{p_0(y,a)}\right]^{\lambda}\right)^{\rho}, \quad -\infty < y < \infty,\ a \geq 0 \tag{4.61}$$

where the parameters α,β are non-negative (see [172, Appendix A]). For specific values of λ, ρ and δ, the parameter α is optimally determined as to satisfy the implicit equation

$$\frac{\displaystyle\int_{-\infty}^{\infty}\int_{0}^{\infty} p_0(y,a) \left(1 + \alpha \left[\frac{p_1(y,a)}{p_0(y,a)}\right]^{\lambda}\right)^{\rho-1} da\, dy}{\displaystyle\int_{-\infty}^{\infty}\int_{0}^{\infty} p_0(y,a) \left(1 + \alpha \left[\frac{p_1(y,a)}{p_0(y,a)}\right]^{\lambda}\right)^{\rho} da\, dy} = 1 - \delta. \tag{4.62}$$

The parameter β is then optimally determined by the following relation (which stems directly from (4.60) and (4.61)):

$$\beta = \left\{\int_{-\infty}^{\infty}\int_{0}^{\infty} p_0(y,a) \cdot \left(1 + \alpha \left[\frac{p_1(y,a)}{p_0(y,a)}\right]^{\lambda}\right)^{\rho} da\, dy\right\}^{-1}. \tag{4.63}$$

We note here that the left hand side of (4.62) is a decreasing function of the non-negative parameter α, and it also admits every value between zero and unity (corresponding to $\alpha \to \infty$ and $\alpha = 0$, respectively). Therefore, the existence and uniqueness of a solution α for (4.62) is assured for any δ (as $0 < \delta < 1$), and this solution can be determined numerically (e.g., by the bisection method).

We observe from our discussion so far that the minimization of the above upper bound on the block error probability (which is based on the minimization of the upper bound on $P_{\mathrm{e}|0}(d)$ in (4.59) and the calculation of the union bound in (4.45)) involves a numerical minimization of (4.59) over the parameters λ and ρ (where $\lambda \geq 0$ and $0 < \rho \leq 1$), for every particular subcode $\mathcal{C}_d$ ($d = 0,1,\ldots,N$). The optimal values of

α and β which are related to the optimal tilting measure ψ in (4.61) are numerically evaluated from (4.62) and (4.63) as a function of the two optimized parameters λ and ρ. This minimization is performed separately for every subcode (where the number of the subcodes $\mathcal{C}_d$ doesn't exceed the length N of the linear block code $\mathcal{C}$, and clearly we are interested only on the subcodes $\mathcal{C}_d$ for which $S_d > 0$, as otherwise $P_{\mathrm{e}|0}(d) = 0$).

Suppose we wish to calculate an upper bound on the block error probability for an ensemble of linear block codes whose *average* distance spectrum is calculable. By invoking the Jensen inequality to the right hand side of (4.59), then $E[(S_d)^\rho] \leq (E[S_d])^\rho$, where $0 \leq \rho \leq 1$. Hence, for ensembles of codes, the upper bound (4.59) therefore stays valid by replacing the distance spectrum in the right hand side of (4.59) with the statistical expectation of the distance spectrum (as was first noted in [62]).

4.4.3.2 Exponential tilting measure

In [173], a sub-optimal selection for ψ in (4.58) is suggested which in fact is motivated by the Duman and Salehi (first version) bound [62]. An exponential tilting measure is also applied in [173] to the fading sample a (treated as a measurement), which gives rise to the exponential tilting measure

$$\psi(y,a) = \frac{\sqrt{\frac{\alpha}{2\pi}} \exp\left[-\frac{\alpha}{2}\left(y - au\sqrt{\frac{2E_{\mathrm{s}}}{N_0}}\right)^2 - \frac{\alpha v^2 a^2 E_{\mathrm{s}}}{N_0}\right] p(a)}{\displaystyle\int_0^{+\infty} p(a) \exp\left(-\frac{\alpha v^2 a^2 E_{\mathrm{s}}}{N_0}\right) da} \qquad (4.64)$$

where $\alpha \geq 0$, $-\infty < u < +\infty$, $-\infty < v < +\infty$, and $p(a)$ designates the probability density function of the independent fading samples. That yields a closed form upper bound in (4.59), which reads (see [173])

$$P_{\mathrm{e}|0}(d) \leq (S_d)^\rho \, \alpha^{\frac{-N(1-\rho)}{2}} \left(\alpha - \frac{\alpha - 1}{\rho}\right)^{-\frac{N\rho}{2}} \cdot \left(\frac{1}{1+t}\right)^{N(1-\rho)} \left(\frac{1}{1+\varepsilon}\right)^{d\rho} \left(\frac{1}{1+v}\right)^{(N-d)\rho} \qquad (4.65)$$

where

$$t = \frac{\alpha \nu^2 E_{\mathrm{s}}}{N_0} \tag{4.66a}$$

$$v = \frac{E_{\mathrm{s}}}{N_0} \left[\alpha(u^2 + \nu^2)\left(1 - \frac{1}{\rho}\right) + \frac{1}{\rho} - \frac{\left(\alpha u - \frac{\alpha u - 1}{\rho}\right)^2}{\alpha - \frac{\alpha - 1}{\rho}} \right] \tag{4.66b}$$

$$\varepsilon = v + \frac{E_{\mathrm{s}}}{N_0} \, \frac{\left(\alpha u - \frac{\alpha u - 1}{\rho}\right)^2 - \left(\alpha u - \frac{\alpha u - 1}{\rho} - 2\lambda\right)^2}{\alpha - \frac{\alpha - 1}{\rho}} \,. \tag{4.66c}$$

This bound is in fact equivalent to the Divsalar and Biglieri bound [52] which has been derived via a geometric extension of the associated decision region in Divsalar bound [50] (by rotating the displaced sphere region in [50]). The bound (4.64)–(4.66) also yields to an extension of the Viterbi and Viterbi (first version) bound for fully interleaved fading channels and perfect CSI at the receiver (see [173]), by setting

$$\alpha = 1\,, \quad u = 1 - \xi\rho\,, \quad \nu = \sqrt{1 - (1 - \xi\rho)^2}\,, \quad \lambda = \frac{1 + \xi\,(1 - \rho)}{2}\,. \tag{4.67}$$

In [173], the DS2 bound associated with the exponential tilting measure (4.64) is also applied to the fully interleaved Rician fading channel. As a special case, it yields the Duman and Salehi (first version) bound for a binary-input AWGN channel, where the Rician parameter K standing for the power ratio of the direct and the diffused received paths, goes to infinity. The bounds are depicted in Fig. 4.5 for the ‘repeat and accumulate’ (RA) codes (an ensemble of turbo-like codes which was introduced by Divsalar, Jin and McEliece [54]) operating over a fully interleaved Rayleigh fading channel. The penalty for the gradual specialization of the DS2 bound by constraining the selection ψ in (4.58) is explicitly indicated.

4.5 Gallager-type bounds for the mismatched decoding regime

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

$$Q_N(\underline{y}|\underline{x}^j) > Q_N(\underline{y}|\underline{x}^i) \ , \quad \forall\, i \neq j \ , \quad i = 1, 2, \ldots, M \ , \tag{4.68}$$

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

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

4.5.1 The mismatched Duman and Salehi bound

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

$$P_{\mathrm{e}|m} \leq \sum_{\underline{y}} p_N(\underline{y}|\underline{x}^m) \left(\sum_{m' \neq m} \left(\frac{Q_N(\underline{y}|\underline{x}^{m'})}{Q_N(\underline{y}|\underline{x}^m)} \right)^{\lambda} \right)^{\rho} , \quad \lambda, \rho \geq 0. \tag{4.69}$$

By invoking the Duman and Salehi bounding technique as described in section 4.2.2, then starting from (4.69) yields in parallel to (4.8),

$$P_{\mathrm{e}|m} \leq \left(\sum_{m' \neq m} \sum_{\underline{y}} p_N(\underline{y}|\underline{x}^m)^{\frac{1}{\rho}} \, \psi_N^m(\underline{y})^{1-\frac{1}{\rho}} \left(\frac{Q_N(\underline{y}|\underline{x}^{m'})}{Q_N(\underline{y}|\underline{x}^m)} \right)^{\lambda} \right)^{\rho} ,$$
$$\lambda \geq 0, \ \ 0 \leq \rho \leq 1 \tag{4.70}$$

where as in Section 4.2.2, $\psi_N^m(\underline{y})$ is the normalized Duman and Salehi tilting measure. In parallel to (4.10), the DS2 bound with the unnormalized tilting measure reads

$$P_{\mathrm{e}|m} \leq \left(\sum_{\underline{y}} G_N^m(\underline{y}) \, p_N(\underline{y}|\underline{x}^m) \right)^{1-\rho}$$
$$\cdot \left(\sum_{m' \neq m} \sum_{\underline{y}} p_N(\underline{y}|\underline{x}^m) \, G_N^m(\underline{y})^{1-\frac{1}{\rho}} \left(\frac{Q_N(\underline{y}|\underline{x}^{m'})}{Q_N(\underline{y}|\underline{x}^m)} \right)^{\lambda} \right)^{\rho} ,$$
$$0 \leq \rho \leq 1, \ \ \lambda \geq 0 \ . \tag{4.71}$$

The advantage of this bound, as compared to the original Gallager mismatched bound (4.69), is as in the matched case: The possible utilization for specific codes and ensembles. This is demonstrated in Section 4.7.2 for a fully interleaved fading channel with faulty channel state information.

4.5.2 Ensembles of random codes

In this subsection, we examine the bound for a random coding strategy. As in the matched regime, for a general distribution of codewords $q_N(\underline{x})$, the DS2 bound with the optimized tilting measure reconstructs the 1965 Gallager random coding bound. We therefore continue with the mismatched Gallager bound (4.69) and restrict our attention to ensembles of codes where each codeword satisfies the inequality

$$N\varepsilon - \delta \leq \Gamma_N(\underline{x}^j) < N\varepsilon\,, \quad \delta > 0\,, \quad j = 1, 2, \ldots, M \tag{4.72}$$

where $\Gamma_N(\underline{x}^j)$ stands for an arbitrary cost function involved in the transmission of the codeword $\underline{x}^j$ and ε is a positive constant, both to be specified. For a codebook satisfying the inequality (4.72) (for every codeword), we further loosen the bound (4.69) letting

$$P_{\mathrm{e}|m} \leq \sum_{\underline{y}} p_N(\underline{y}|\underline{x}^m) \left[\sum_{m' \neq m} \left(\frac{Q_N(\underline{y}|\underline{x}^{m'})}{Q_N(\underline{y}|\underline{x}^m)} \right)^{\lambda} \left(\frac{e^{\Gamma_N(\underline{x}^{m'})}}{e^{\Gamma_N(\underline{x}^m)}} \right) e^{\delta} \right]^{\rho}, \\ \lambda, \rho \geq 0\,. \tag{4.73}$$

This is since for every value of j $(j = 1, 2, \ldots, M)$, the inequality $\frac{e^{\Gamma_N(\underline{x}^{m'})}}{e^{\Gamma_N(\underline{x}^m)}} e^{\delta} > 1$ holds for any codebook satisfying (4.72). Now, we select randomly and independently each codeword in such a codebook by the probability law $\alpha_N(\underline{x})$. The upper bound on the decoding error probability over this ensemble equals

$$P_{\mathrm{e}} \leq \sum_{\underline{y}} \sum_{\underline{x}^m} \alpha_N(\underline{x}^m) p_N(\underline{y}|\underline{x}^m) \sum_{\underline{x}^1} \cdots \sum_{\underline{x}^{m-1}} \sum_{\underline{x}^{m+1}} \cdots \sum_{\underline{x}^M} \left\{ \prod_{\substack{i=1 \\ i \neq m}}^{M} \alpha_N(\underline{x}^i) \right. \\ \left. \cdot \left(\sum_{j \neq m} \left(\frac{Q_N(\underline{y}|\underline{x}^j)}{Q_N(\underline{y}|\underline{x}^m)} \right)^{\lambda} \frac{e^{\Gamma_N(\underline{x}^j)}}{e^{\Gamma_N(\underline{x}^m)}} e^{\delta} \right)^{\rho} \right\} \tag{4.74}$$

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

As indicated in (4.72), the measure $\alpha_N(\underline{x})$ satisfies the condition

$$\alpha_N(\underline{x}) \equiv 0 \qquad \forall\, \underline{x} \,:\, \Gamma_N(\underline{x}) \notin [N\varepsilon - \delta, N\varepsilon)\ . \tag{4.75}$$

In principle, the parameter δ introduced in (4.72) may depend on N. Invoking the Jensen inequality in (4.74) yields the following upper bound on the decoding error probability:

$$P_{\mathrm{e}} \leq (M-1)^{\rho}\, e^{\delta\rho} \sum_{\underline{y}} \sum_{\underline{x}} \alpha_N(\underline{x})\, p_N(\underline{y}|\underline{x}) \cdot \left\{ \sum_{\underline{x}'} \alpha_N(\underline{x}') \left(\frac{Q_N(\underline{y}|\underline{x}')}{Q_N(\underline{y}|\underline{x})} \right)^{\lambda} \frac{e^{\Gamma_N(\underline{x}')}}{e^{\Gamma_N(\underline{x})}} \right\}^{\rho} \tag{4.76}$$

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

$$\alpha_N(\underline{x}) = \frac{q_N(\underline{x})}{\mu_\alpha} \tag{4.77}$$

where based on (4.72)

$$\mu_\alpha = \mu_\alpha(\Gamma, \delta, \varepsilon) = \sum_{\underline{x}:\Gamma_N(\underline{x}) \in [N\varepsilon-\delta, N\varepsilon)} q_N(\underline{x})\ . \tag{4.78}$$

The substitution of (4.77) into (4.76) gives

$$P_{\mathrm{e}} \leq \frac{(M-1)^{\rho}\, e^{\delta\rho}}{(\mu_\alpha)^{1+\rho}} \sum_{\underline{y}} \sum_{\underline{x}} q_N(\underline{x})\, p_N(\underline{y}|\underline{x}) \cdot \left\{ \sum_{\underline{x}'} q_N(\underline{x}') \left(\frac{Q_N(\underline{y}|\underline{x}')}{Q_N(\underline{y}|\underline{x})} \right)^{\lambda} \frac{e^{\Gamma_N(\underline{x}')}}{e^{\Gamma_N(\underline{x})}} \right\}^{\rho} \tag{4.79}$$

where $\lambda \geq 0$, $0 \leq \rho \leq 1$ and $\delta > 0$. We may further loosen the bound (4.79) by replacing $\frac{e^{\delta\rho}}{(\mu_\alpha)^{1+\rho}}$ by $\frac{e^{\delta}}{\mu_\alpha^2}$, as $\delta > 0$, $0 \leq \rho \leq 1$ and $\mu_\alpha \leq 1$.

This yields

$$P_{\mathrm{e}} \leq \frac{e^{\delta} M^{\rho}}{\mu_{\alpha}^{2}} \sum_{\underline{y}} \sum_{\underline{x}} q_N(\underline{x}) \, p_N(\underline{y}|\underline{x}) \cdot \left\{ \sum_{\underline{x}'} q_N(\underline{x}') \left(\frac{Q_N(\underline{y}|\underline{x}')}{Q_N(\underline{y}|\underline{x})} \right)^{\lambda} \frac{e^{\Gamma_N(\underline{x}')}}{e^{\Gamma_N(\underline{x})}} \right\}^{\rho} \tag{4.80}$$

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

$$p_N(\underline{y}|\underline{x}) = \prod_{k=1}^{N} p(y_k|x_k), \; Q_N(\underline{y}|\underline{x}) = \prod_{k=1}^{N} Q(y_k|x_k) \tag{4.81}$$

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

$$q_N(\underline{x}) = \prod_{k=1}^{N} q(x_k). \tag{4.82}$$

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

$$\Gamma_N(\underline{x}) = \sum_{k=1}^{N} \gamma(x_k). \tag{4.83}$$

Substituting (4.81)-(4.83) into (4.80) then yields the single letter expression

$$P_{\mathrm{e}} \leq \frac{e^{\delta}}{\mu_{\alpha}^{2}} e^{\rho R N} \left[\sum_{y} \sum_{x} q(x) \, p(y|x) \left(\sum_{x'} q(x') \left(\frac{Q(y|x')}{Q(y|x)} \right)^{\lambda} \frac{e^{\gamma(x')}}{e^{\gamma(x)}} \right)^{\rho} \right]^{N} \tag{4.84}$$

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

$$P_{\mathrm{e}} \leq \frac{e^{\delta}}{\mu_{\alpha}^{2}} \cdot e^{-N[E_0(q,\gamma,\rho,\lambda) - \rho R]} \tag{4.85}$$

where

$$E_0(q,\gamma,\rho,\lambda) = - \ln \left[\sum_{y} \sum_{x} q(x) \, p(y|x) \left(\sum_{x'} q(x') \left(\frac{Q(y|x')}{Q(y|x)} \right)^{\lambda} \frac{e^{\gamma(x')}}{e^{\gamma(x)}} \right)^{\rho} \right]. \tag{4.86}$$

Before turning to treat the exponent in (4.85), let us estimate μ_α for fixed (N-independent) δ. Based on (4.78) and (4.83)

$$\mu_\alpha = \Pr\left(-\delta \le \sum_{k=1}^{N} \gamma(x_k) - N\varepsilon < 0\right) \tag{4.87}$$

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

$$\varepsilon = \sum_x q(x)\,\gamma(x) \tag{4.88}$$

for a fixed function γ. Similarly to Gallager [83], one can estimate μ_α for $N \to \infty$, yielding

$$\lim_{N\to\infty} \sqrt{N}\,\mu_\alpha = \frac{\delta}{\sqrt{2\pi\sigma^2}} \tag{4.89}$$

where

$$\sigma^2 = \sum_x q(x)\,\gamma(x)^2 - \left(\sum_x q(x)\,\gamma(x)\right)^2. \tag{4.90}$$

This is directly found based on (4.87) by the central limit property as follows:

$$\begin{aligned} \mu_\alpha &= \Pr\left(\frac{-\delta}{\sqrt{N}} \le \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \gamma(x_k) - \sqrt{N}\varepsilon < 0\right) \\ &\xrightarrow[N\to\infty]{} \Pr\left(\frac{-\delta}{\sqrt{N}} \le N(0,\sigma^2) < 0\right) \\ &\underset{\frac{\delta}{\sqrt{N}} \ll 1}{\approx} \frac{\delta}{\sqrt{N}}\,\frac{1}{\sqrt{2\pi\sigma^2}}\,. \end{aligned} \tag{4.91}$$

The pre-exponent factor in (4.85) can be optimized over $\delta \ge 0$ to yield

$$\min_{\delta>0} \lim_{N\to\infty} \frac{1}{N}\,\frac{e^\delta}{\mu_\alpha^2} = \min_{\delta>0} (2\pi\sigma^2) \cdot \frac{e^\delta}{\delta^2} = (2\pi\sigma^2) \cdot \frac{e^2}{4} \tag{4.92}$$

which demonstrates that the behavior of the pre-exponent in (4.85) is asymptotically proportional to N (with $N \to \infty$). Thus for finite

ε (4.88) and σ^2 (4.90), the pre-exponent in (4.85) has no exponential implications as it behaves asymptotically like a $\frac{\ln(N)}{N}$ term in the exponent. For a fixed input distribution $q(x)$, we then attempt to maximize the error exponent

$$E(R,q) = \sup_{\substack{0 \le \rho \le 1 \\ \lambda \ge 0 \\ \gamma(x)}} \Big(E_0(q,\gamma,\rho,\lambda) - \rho R \Big) \tag{4.93}$$

where $\gamma(x)$ is a real function yielding finite ε and σ^2 in (4.88) and (4.90) respectively. Consider the rate equation

$$\begin{aligned} R &= \frac{\partial}{\partial \rho} \Big(E_0(q,\gamma,\rho,\lambda) \Big) \\ &= e^{E_0(q,\gamma,\rho,\lambda)} \\ &\quad \cdot \left\{ -\sum_y \sum_x q(x)\, p(y|x) \left(\sum_{x'} q(x') \left(\frac{Q(y|x')}{Q(y|x)} \right)^{\lambda} \frac{e^{\gamma(x')}}{e^{\gamma(x)}} \right)^{\rho} \right. \\ &\quad \left. \cdot \ln \left[\frac{\sum_{x'} q(x')\, Q^{\lambda}(y|x')\, e^{\gamma(x')}}{Q^{\lambda}(y|x)\, e^{\gamma(x)}} \right] \right\} \end{aligned} \tag{4.94}$$

where the last equality results from (4.86). The maximal rate R_H is calculated by substituting $\rho = 0$ in (4.94), yielding

$$R_H = \sup_{\gamma(x)} \max_{\lambda \ge 0} \sum_y \sum_x q(x)\, p(y|x) \ln \left(\frac{Q^{\lambda}(y|x)\, e^{\gamma(x)}}{\sum_{x'} q(x')\, Q^{\lambda}(y|x')\, e^{\gamma(x')}} \right) \tag{4.95}$$

This, in fact, equals what is known as the Csiszár-Körner-Hui lower bound on the mismatched capacity, since it yields the dual representation in the terminology of [85]. It is rather straightforwardly verified that the error exponent $E(R)$ in (4.93) is positive for $R < R_H$. The input distribution $q(x)$ can now be chosen to maximize R_H, in the usual sense [85], [132]. In the case of continuous input and output alphabets, the relevant sums should be replaced by integrals and it should be verified that the optimized $\gamma(x)$ and $q(x)$ yield finite (though arbitrary) ε and σ in (4.88) and (4.90) respectively.

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

$$R = \sup_{\lambda \geq 0} E_0(q, \gamma = 0\,, \rho = 1, \lambda) \tag{4.96}$$

yields the generalized cutoff rate [106], which cannot be further improved by optimizing over the real function $\gamma(x)$. In terms of ensembles, it is worth emphasizing that the upper bound (4.85) resulted by restricting our attention to ensembles of codewords satisfying (4.72). This restriction is necessary, as if a purely random ensemble is attempted, then it was concluded in [85] that the results associated with $\gamma(x) = 0$ cannot be surpassed. The maximal rate then corresponds to the generalized mutual information (GMI) [132].

In [153], the 1965 Gallager bounding technique was applied to derive upper bounds for i.i.d. random and fixed composition codes, operating over memoryless channels with finite input alphabets and arbitrary output alphabets, and whose decoding metric is matched to the channel. The metric which is at the beginning general, was then optimized to yield an equivalent metric to ML decoding, and yet the results in [153] reproduced the error exponent for fixed composition codes [41]. Here we deal with a mismatched metric and the code ensemble is restricted by introducing a generalized energy constraint (4.72), which is subject to optimization.

4.5.3 Ensembles of structured codes

We apply here the generalization of the DS2 bound for the mismatched decoding regime (4.69) to ensembles of structured codes. We assume that the transition probabilities of the channel $p_N(\underline{y}|\underline{x})$ and the mismatched metric $Q_N(\underline{y}|\underline{x})$ are MBIOS (i.e., memoryless, binary-input and output-symmetric).

The optimization of the normalized Duman and Salehi tilting measure $\psi_N^m(\underline{y})$ in (4.70) is restricted here to the case where $\psi_N^0(\underline{y})$ can be expressed in the product form $\psi_N^0(\underline{y}) = \prod_{i=1}^N \psi(y_i)$. In that case, the partitioning of the code (or ensemble of codes) to constant Hamming weight subcodes yields (4.45), where

$$P_{\mathrm{e}|0}(d) \le (S_d)^\rho \left(\sum_y g_1(y)\, \psi(y)^{1-\frac{1}{\rho}}\right)^{(N-d)\rho} \left(\sum_y g_2(y)\, \psi(y)^{1-\frac{1}{\rho}}\right)^{d\rho} \tag{4.97}$$

and

$$g_1(y) = p(y|0)^{\frac{1}{\rho}}\ , \quad g_2(y) = p(y|0)^{\frac{1}{\rho}} \left(\frac{Q(y|1)}{Q(y|0)}\right)^\lambda . \tag{4.98}$$

For continuous output channels, the sums in (4.97) should be replaced by integrals. With the aid of calculus of variations, the optimal normalized Duman and Salehi tilting measure ψ, in terms of minimizing the upper bound (4.97)-(4.98), admits the form

$$\psi(y) = \frac{p(y|0)\left(1+\alpha\left(\frac{Q(y|1)}{Q(y|0)}\right)^\lambda\right)^\rho}{\sum_y p(y|0)\left(1+\alpha\left(\frac{Q(y|1)}{Q(y|0)}\right)^\lambda\right)^\rho}\ , \quad \lambda \ge 0\,,\ 0 \le \rho \le 1\ . \tag{4.99}$$

As demonstrated in [183, Appendix B], the parameter α is optimally determined by a numerical solution of the equation

$$\frac{\sum_y p(y|0)\left(1+\alpha\left(\frac{Q(y|1)}{Q(y|0)}\right)^\lambda\right)^{\rho-1}}{\sum_y p(y|0)\left(1+\alpha\left(\frac{Q(y|1)}{Q(y|0)}\right)^\lambda\right)^\rho} = 1-\delta \tag{4.100}$$

where $\delta \equiv \frac{d}{N}$ is the normalized Hamming weight of the N-length codewords possessing a Hamming weight d. The existence and uniqueness of a solution α in (4.100) is proved in [183, Appendix B].

From the discussion above, the upper bound for the mismatched decoding regime in (4.45), (4.97)–(4.100) involves numerical optimizations over the two parameters $\lambda \ge 0$, $0 \le \rho \le 1$, for at most the N constant Hamming weight subcodes of the considered ensemble of codes. The necessary information on the ensemble of codes comprises the average distance spectrum (or even a tight upper bound on the ensemble distance spectrum), so that more refined information on the algebraic structure of the codes is not required.

4.6 Gallager-type bounds for parallel channels

The error performance analysis in the case where each codeword is partitioned and each part is transmitted over one of independent channels in parallel is of interest. Code partitioning is employed in transmission over block-fading channels (for performance bounds of coded communication systems over block-fading channels, see [70, 219, 227, 228]), incremental redundancy retransmission schemes, cooperative coding, multi-carrier signaling (for performance bounds of coded orthogonal-frequency division multiplexing (OFDM) systems, see [119, 226]), etc.

In his thesis [64], Ebert considered the problem of communicating over parallel discrete time channels, disturbed by an additive Gaussian noise with a total power constraint on the set of channels. 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 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.

In [122], Liu et al. derive upper bounds on the ML decoding error probability of structured ensembles of codes whose transmission takes place over (independent) parallel channels. The analysis in [122] modifies the 1961 Gallager-Fano bound [81, Section 3] and adapts this bounding technique for the communication over parallel channels. As special cases of this modified bound, a generalization of the union-Bhattacharyya bound, the SFB [187], sphere bound, and a combination of the two former bounds are derived for parallel channels. In order to make the calculation of the weight enumerators feasible when taking into account the partitioning of codewords and their transmission over parallel channels, the authors study the case where the bits of the codewords are randomly assigned to J parallel channels, where the fraction of bits transmitted over each channel is determined in advance. The upper bounds on the ML decoding error probability are applied to ensembles of codes defined on graphs (e.g., uniformly interleaved repeat-accumulate codes and turbo codes). The comparison between upper bounds under ML decoding and computer simulations of the performance of such ensembles under iterative decoding shows a good match in several cases. For a given ensemble of codes and a given codeword-symbol to channel assignment rule, a reliable channel region is defined as the closure of the set of parallel-channel transition probabilities for which the decoding error probability vanishes as the codeword length goes to infinity. The upper bounds on the block error probability derived in [122] enable to derive achievable regions for ensuring reliable communications under ML decoding.

Tightened Gallager bounds for independent parallel MBIOS channels were recently derived in [165, 166], and were exemplified for various ensembles of turbo-like codes (see Section 4.7.5). Performance bounds for dependent parallel channels are considered in [12].

4.7 Some applications of the Gallager-type bounds

In Section 4.4, a large class of efficient bounds (or their Chernoff versions) was demonstrated to be a special case of the generalized DS2 bound. Implications and applications of these observations are pointed out here, including the fully interleaved fading channel, resorting to either matched or mismatched decoding. The proposed approach can be also generalized to geometrically uniform non-binary codes, finite state channels, bit interleaved coded modulation systems, and it can be also used for the derivation of upper bounds on the conditional decoding error probability.

4.7.1 AWGN channels

We apply here some variants of the Gallager bounds and other reported bounds to the ensemble of un-punctured turbo codes of rate $\frac{1}{3}$ with a uniform interleaver of length 1000 and two recursive systematic convolutional (RSC) component codes whose generators are $(1, \frac{21}{37})$ in octal form (see Fig. 4.2). The considered ensemble of codes is also

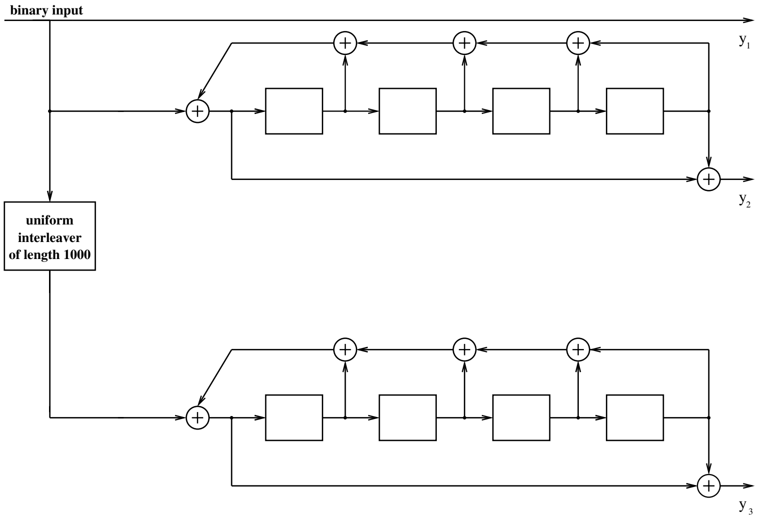

Fig. 4.2 Block diagram of an ensemble of uniformly interleaved turbo codes of rate-$\frac{1}{3}$ and interleaver length 1000. The generator of the two identical component codes (16 states, recursive systematic convolutional (RSC) codes) is $\left[1, \frac{1+D^4}{1+D+D^2+D^3+D^4}\right]$. A termination to the all-zero state is assumed at the end of each block.

terminated to the all-zero state at the end of each frame by additional four bits (having thus an overall of 3012 coded bits). The following upper bounds on the bit error probability are depicted in Fig. 4.3: The original TSB of Poltyrev ([152, 170]) (differing from the loosened Chernoff version bound derived in [50]), the Engdahl and Zigangirov bound [67] (the non-Chernoff version in Section 4.4.2.5 here), Duman and Salehi bounds [62, 60] (Sections 4.2.2 and 4.4.2.1 here), Divsalar bound [50] (Section 4.4.2.4 here), Viterbi and Viterbi bounds [209, 210] (Sections 4.4.2.2 and 4.4.2.3 here), and finally the union bound. It is demonstrated in Fig. 4.3, that the difference between the two versions of the Duman and Salehi bounds for the binary-input AWGN channel is very small (about 0.01 dB) and that also the Duman and Salehi (first version) bound coincides with Divsalar bound. The first observation is consistent with [60] and the second observation verifies numerically the fact that Divsalar bound is indeed a closed form of Duman and Salehi (first version) bound [62], as was first indicated in [50] (the negligible difference between the two bounds depicted in Fig. 4.3 is attributed to the numerical optimizations associated with the calculation of the first version of the Duman and Salehi bound (4.51)). The TSB [152] and the Engdahl and Zigangirov bound [67] were derived without invoking the Chernoff bounding technique, explaining therefore their advantage over Duman and Salehi bounds and some other related bounds for block codes of moderate block-length, as the latter bounds rely on the Chernoff bounding technique (see Fig. 4.3). However, as indicated in [50] (and Section 4.4.2), the loosening of the TSB which turns it into a

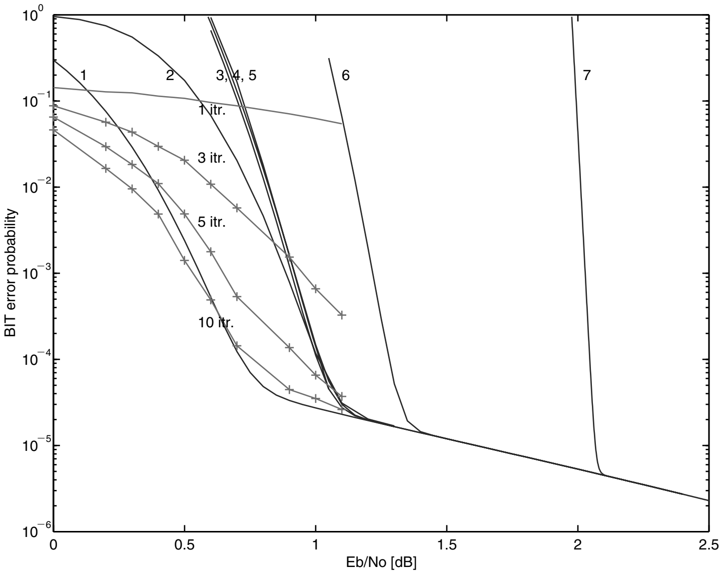

Fig. 4.3 A comparison between upper bounds on the bit error probability with ML decoding. The bounds refer to the ensemble of turbo codes in Fig. 4.2, operating over a binary-input AWGN channel.
1. Tangential-sphere bound (TSB) ([152], [170]).
2. The Engdahl and Zigangirov bound [67].
3. The DS2 (second version of the Duman and Salehi) bound [60].
4. Divsalar bound [50].
5. Duman and Salehi (first version) bound [62].
6. Viterbi & Viterbi (second version) bound [210].
7. Union bound in Q-form.
The bounds 2–6 are combined with the union bound in its Q-form for every constant Hamming-weight subcode of the considered ensemble. We also plot here simulation results of Log-MAP iterative decoding with 1, 3, 5 and 10 iterations.

particular case of the DS2 bound, does not carry any implications on its associated error exponent for the asymptotically infinite block length.

4.7.2 Fully interleaved fading channels

We apply here various variants of the generalized DS2 bound to the fully interleaved Rayleigh fading channel with perfect channel state information (see Section 4.4.3). These bounds are evaluated for the ensemble of rate $\frac{1}{3}$ uniformly interleaved turbo codes depicted in Fig. 4.2. The optimization of the generalized DS2 bound (combined with the tight version of the union bound) is demonstrated as the tightest reported bound (see also [173],[172]). It also approximately replicates here the performance of these turbo codes when iteratively decoded with 10 iterations of the Log-MAP iterative decoding algorithm (see Fig. 4.4). These bounds

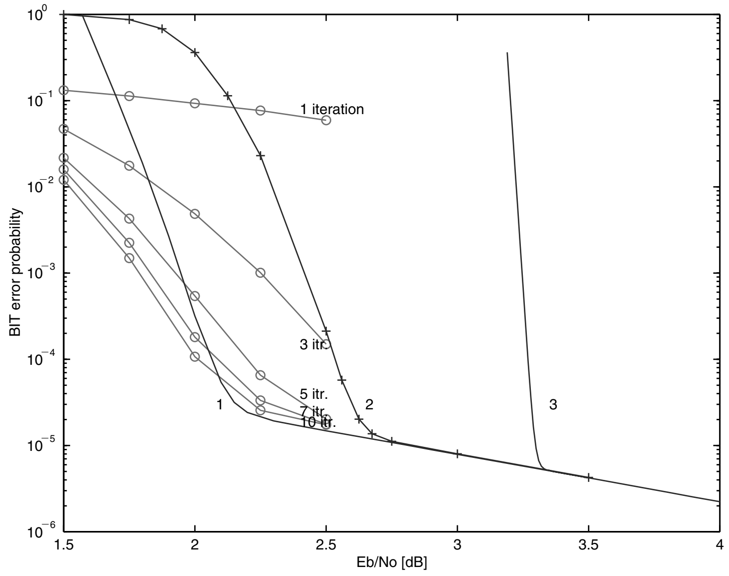

Fig. 4.4 A comparison between upper bounds on the bit error probability with ML decoding. The bounds refer to the ensemble of turbo codes in Fig. 4.2, operating over a fully interleaved Rayleigh fading channel with perfect channel state information.
1. The generalized DS2 bound [172] combined with the union bound in its tight form.
2. The generalization of the Engdahl and Zigangirov bound [172].
3. The union bound in its tight form.
These upper bounds are also compared to computer simulation results of the Log-MAP iterative decoding algorithm with up to 10 iterations.

are also evaluated for the ensemble of rate-$\frac{1}{4}$ uniformly interleaved RA codes [54], with a block length of $N = 4096$ (see upper plot of Fig. 2.1 and Fig. 4.5). In [172, 173], these bounds were applied to some ensembles of efficient codes (turbo [24], Gallager-LDPC [81] and RA codes [54]). For moderate values of energy per bit to spectral noise density $\left(\frac{E_b}{N_0}\right)$, the optimized DS2 upper bound (under ML decoding) falls below the computer simulation results for the sum-product iterative decoding algorithm (see Fig. 4.5), demonstrating the mild sub-optimality of iterative decoding.

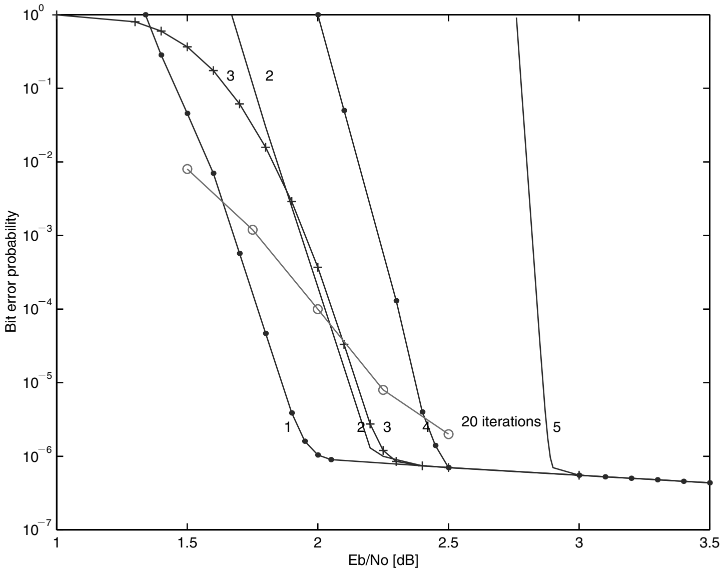

Fig. 4.5 Comparison between upper bounds on the bit error probability with ML decoding. The bounds refer to the RA codes in the upper plot of Fig. 2.1, operating over a fully interleaved Rayleigh fading channel with perfect channel state information.
1. The generalized DS2 bound [172].
2. Divsalar and Biglieri bound [52].
3. The generalization of the Engdahl and Zigangirov bound [172].
4. The generalization of Viterbi and Viterbi bound [173].
5. The tight form of the union bound.
The upper bounds 1–4 are combined with the tight version of the union bound for every constant Hamming-weight subcode of the considered ensemble of codes. The bounds are also compared with computer simulation results of the sum-product iterative decoding algorithm with 20 iterations.

4.7.3 Thresholds of codes

Let $\mathcal{C}$ be an ensemble of binary, linear block codes of block-length N and rate R. Consider the case where the transmission takes place over an AWGN channel. The threshold is defined as the infimum over the values of $\frac{E_{\mathrm{b}}}{N_0}$ ensuring a decoding error probability which asymptotically decays to zero as the block length tends to infinity. Obviously, the threshold depends on four parameters: The channel model, the applied decoding algorithm and the considered code (or ensemble of codes), and whether we require a vanishing block error probability or a vanishing bit error probability. In the following, we consider soft-decision ML decoding, and require the stronger constraint of vanishing block error probability.

In Table 4.1, bounds on the thresholds are exemplified for various ensembles of regular LDPC codes and uniformly interleaved RA codes. The bounds refer to ML decoding where transmission takes place over a binary-input AWGN channel. The bounds on the thresholds are compared to their exact values under iterative message-passing decoding (using the density evolution technique), and are also compared with the Shannon capacity limit as a reference. It is shown that as the code rate is increased, the gap between the bounds on the thresholds which follow from the DS2 and Divsalar bounds becomes more pronounced (where the bound of Divsalar was introduced in [50] and Section 3.2.4 here).

Table 4.1 Upper bounds on the thresholds of $\frac{E_{\mathrm{b}}}{N_0}$ for ensembles of RA and LDPC codes. The bounds refer to ML decoding where transmission takes place over a binary-input AWGN channel. The bounds are compared to the exact values of the thresholds under iterative message-passing decoding (using the density evolution technique for the sum-product decoding algorithm [156]), and are also compared as a reference with the Shannon capacity limit.

Ensemble	**Rate**	**Divsalar bound**	**DS2 bound**	**Density evolution**	**Capacity**
RA($q = 3$)	$\frac{1}{3}$	0.792 dB	0.752 dB	0.479 dB	−0.495 dB
RA($q = 4$)	$\frac{1}{4}$	−0.052 dB	−0.068 dB	0.106 dB	−0.794 dB
RA($q = 5$)	$\frac{1}{5}$	−0.480 dB	−0.488 dB	0.044 dB	−0.936 dB
LDPC(3, 6)	$\frac{1}{2}$	0.793 dB	0.679 dB	1.110 dB	0.187 dB
LDPC(4, 6)	$\frac{1}{3}$	−0.381 dB	−0.419 dB	1.674 dB	−0.495 dB
LDPC(3, 4)	$\frac{1}{4}$	−0.488 dB	−0.508 dB	1.003 dB	−0.794 dB

The bounds also indicate on the sub-optimality of iterative message-passing decoding (e.g., for the ensemble of (4,6) LDPC codes, there is a gap of more than 2 dB between the thresholds under ML and iterative decoding algorithms).

Upper bounds on the thresholds of RA [54] and Gallager-LDPC codes [81] under “typical pairs” decoding algorithm, were derived by McEliece et al. [4] for the binary symmetric channel (BSC) and for the binary-input AWGN channel. Exact thresholds for the BSC under a specific iterative decoding algorithm were derived by Bazzi et al. [15].

Based on the optimized DS2 bound, an upper bound on the thresholds which are associated with ML decoding were numerically calculated for the fully interleaved fading channels with perfect channel state information (see Section 3.3.1). As predicted, these thresholds meet the ultimate Shannon capacity limit for the ensemble of fully random block codes of rate R. Thresholds for the ensembles of RA and (j,k) Gallager-LDPC codes are depicted in Table 4.2. These thresholds refer to the fully interleaved Rayleigh fading channel with perfect channel state information and a maximum ratio combining space diversity of order four. As expected, due to the simple structure of RA codes, the calculated thresholds for the ensemble of RA codes are worse than the corresponding thresholds for the ensembles of Gallager-LDPC codes.

Table 4.2 Upper bounds on the $\frac{E_\mathrm{b}}{N_0}$ – thresholds for ensembles of codes operating over a fully interleaved Rayleigh fading channel with space diversity (based on the maximum ratio combining principle) of order four, and perfect channel state information at the receiver.

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

In general, the calculation of the upper bounds on these thresholds depends on the asymptotic exponent of the distance spectra of the considered codes (or ensembles). Recently, some techniques for the calculation of the asymptotic exponents of the distance spectra of turbo codes and irregular LDPC codes where derived in [174] and [34], respectively. These techniques can be applied to calculate upper bounds on the $\frac{E_\mathrm{b}}{N_0}$-thresholds for turbo codes and LDPC codes under optimal ML decoding.

4.7.4 Mismatched decoding

We apply here the generalized DS2 bound introduced in Section 4.5.3, to study the robustness of a mismatched decoder that is based on ML decoding with respect to the faulty channel measurements. We examine here a BPSK modulated signal, transmitted through a fully interleaved Rayleigh fading channel. For simplicity, we apply our bounds to the case of perfect phase estimation of the i.i.d. fading samples (in essence reducing the problem to a real channel). We also assume that the estimated and real magnitudes of the Rayleigh fading samples satisfy a joint distribution of two correlated bivariate Rayleigh variables with an average power of unity.

Based on the notation in Section 4.5, we therefore obtain for the MBIOS channel

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

where $a, \hat{a}$ are jointly Rayleigh distributed according to the following *pdf*:

$$p_{a,\hat{a}}(r_1, r_2) = \frac{4r_1 r_2}{1-\rho^*}\, I_0\left(\frac{2\sqrt{\rho^*}\, r_1 r_2}{1-\rho^*}\right) \exp\left(-\frac{r_1^2 + r_2^2}{1-\rho^*}\right), \quad r_1, r_2 \geq 0 \tag{4.101}$$

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

$$\rho^* = \frac{\mathrm{Cov}\,(a^2, \hat{a}^2)}{\sqrt{\mathrm{Var}(a^2)\,\mathrm{Var}(\hat{a}^2)}}\,. \tag{4.102}$$

The integral expressed in (4.101) can be transformed to another integral which is easily calculated with the aid of the Gaussian numerical integration formula (see [183, Appendix C]).

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

$$S_d\, E\left[Q\left(\sqrt{\frac{2RE_\mathrm{b}}{N_0}}\, \frac{\sum_{i=1}^{d} a_i\, \hat{a}_i}{\sqrt{\sum_{i=1}^{d} \hat{a}_i^2}}\right)\right] \tag{4.103}$$

where S_d designates the number of codewords of Hamming weight d, the Q-function is introduced in (2.8) and the notation E stands for the statistical expectation with respect to the i.i.d. Rayleigh fading samples $\{\underline{a}_i\}$ and their Rayleigh distributed estimations $\{\underline{\hat{a}}_i\}$. The expressions in (4.103) (where $0 \leq d \leq N$) are calculated via the Monte-Carlo method, by generating the correlated bivariate Rayleigh random variables with a certain correlation coefficient ρ^* (expressed in (4.102)), based on the algorithm proposed in [194]. The upper bound on the bit error probability based on the generalized DS2 bound in (4.97)–(4.100) is compared in Fig. 4.6 to the improved upper bound that combines the DS2 bound with the tight version of the union bound for every constant Hamming

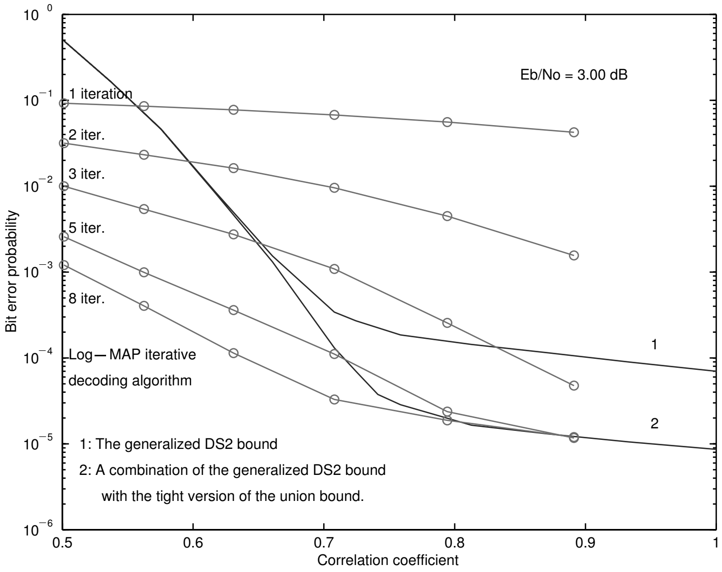

Fig. 4.6 A comparison between two upper bounds on the bit error probability. The bounds refer to the ensemble of turbo codes depicted in Fig. 4.2, operating over a fully interleaved Rayleigh fading channel with mismatched decoding. The generalized DS2 bound is compared with the improved bounding technique which combines the DS2 bound with the tight form of the union bound (for every constant Hamming weight subcode). The bounds are demonstrated for $\frac{E_{\mathrm{b}}}{N_0} = 3$ dB, and are depicted as a function of the correlation coefficient (in the range $\frac{1}{2}$ to 1) between the squares of the i.i.d. jointly Rayleigh fading samples and their estimates.

weight subcode. The comparison between these two bounds refers to the ensemble of uniformly interleaved turbo codes depicted in Fig. 4.2 where $\frac{E_{\mathrm{b}}}{N_0} = 3$ dB. It is reflected from Fig. 4.6 that the latter bound yields a considerably tighter error floor (the error floor in Figs. 4.6 and 4.7 is observed for a sufficiently high correlation coefficient ρ^*, and it reflects the robustness of the system to faulty measurements of the fading samples). The error floor exhibited by the improved upper bound in Fig. 4.6 also predicts reliably the error floor of the turbo codes associated with the suboptimal and efficient Log-MAP iterative decoding algorithm (based on computer simulations with 10 iterations). Note that this error floor depends on the correlation coefficient rather than the SNR. The upper bounds on the bit error probability that are based on the combination of the generalized DS2 bound in (4.97)–(4.100) and

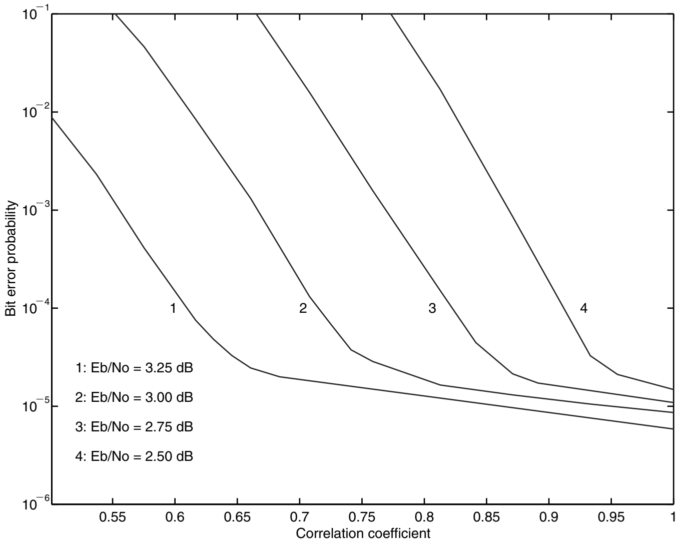

Fig. 4.7 A comparison between upper bounds on the bit error probability for the ensemble of turbo codes depicted in Fig. 4.2. The codes operate over a fully interleaved Rayleigh fading channel with mismatched decoding. The bounds are based on combining the generalized DS2 bound with the tight form of the union bound, and it is applied to every constant Hamming weight subcode. The bounds are plotted for $\frac{E_{\mathrm{b}}}{N_0} = 2.50, 2.75, 3.00$ and 3.25 dB, as a function of the correlation coefficient (in the range $\frac{1}{2}$ to 1) between the squares of the i.i.d. jointly Rayleigh fading samples and their estimates.

the tight version of the union bound (4.103) are depicted in Fig. 4.7 for several fixed values of $\frac{E_{\mathrm{b}}}{N_0}$. These curves are plotted as a function of the correlation coefficient ρ^* between the squares of the i.i.d. jointly Rayleigh fading samples and their estimates. The bounds in Fig. 4.7 refer to the ensemble without puncturing of rate $\frac{1}{3}$ turbo codes depicted in Fig. 4.2. Since for a fully interleaved Rayleigh fading channel with *perfect* side information on the fading samples, the channel cutoff rate corresponds to $\frac{E_{\mathrm{b}}}{N_0} = 3.23$ dB, then according to the upper bounds depicted in Fig. 4.7, the ensemble performance of these turbo codes (associated with the hypothetical ML decoding) is sufficiently robust in case of mismatched decoding, even in a portion of the rate region exceeding the channel cutoff rate.

The proposed upper bounds depicted in Figs. 4.6 and 4.7 were efficiently implemented in Matlab software, indicating their applicability in terms of complexity and the practical running time involved in their calculations.

4.7.5 Parallel channels

In [122], the 1961 Gallager bound was generalized for the case where the communication model forms a parallel concatenation of J statistically independent memoryless binary-input output-symmetric (MBIOS) channels, as shown in Fig. 4.8.

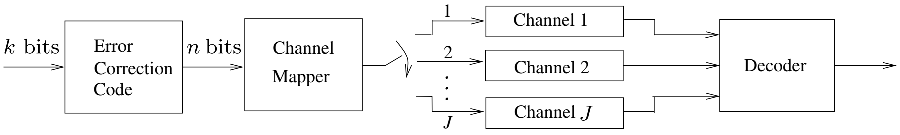

Fig. 4.8 System model of parallel channels. A random mapper is assumed where every bit is assigned to one of the J channels; a bit is assigned to the j^{th} channel independently of the other bits and with probability α_j (where $\sum_{j=1}^{J} \alpha_j = 1$).

Using an error-correcting 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. 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 [122] which then 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 (which is termed the *assignment rate*) the 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 [122] 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.

In [122], Liu et al. derive upper bounds on the ML decoding error probability of structured ensembles of codes whose transmission takes place over (independent) parallel channels. The analysis in [122] modifies the 1961 Gallager-Fano bound from [81, Section 3] (see Section 4.2.3 here) and adapts this bounding technique for the communication over parallel channels. As special cases of this modified bound, a generalization of the union-Bhattacharyya bound, the SFB [187], sphere bound, and a combination of the two former bounds are derived for parallel channels. The upper bounds on the ML decoding error probability are applied to ensembles of codes defined on graphs (e.g., uniformly interleaved repeat-accumulate codes and turbo codes). The comparison in [122] between upper bounds under ML decoding and computer simulations of the performance of such ensembles under iterative decoding shows a good match in several cases. For a given ensemble of codes and a given codeword-symbol to channel assignment rule, a reliable channel region is defined as the closure of the set of parallel-channel transition probabilities for which the decoding error probability vanishes as the codeword length goes to infinity. The upper bounds on the block error probability derived in [122] enable to derive achievable regions for ensuring reliable communications under ML decoding.

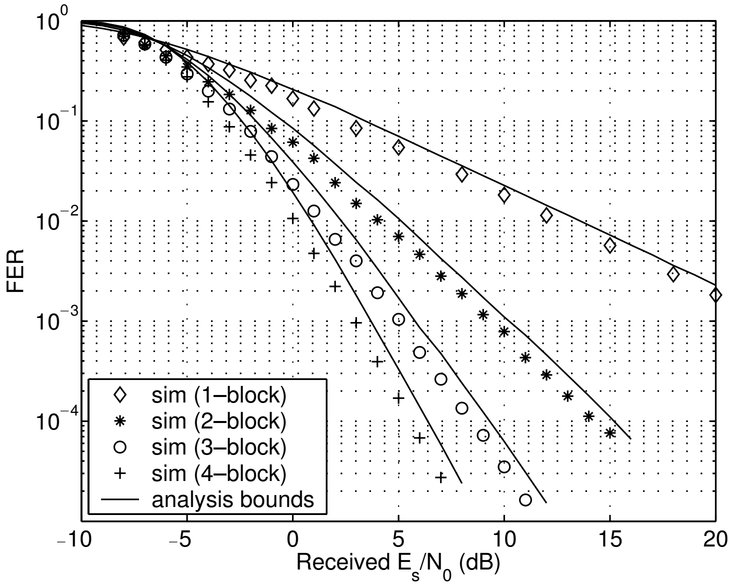

Fig. 4.9 Performance of turbo codes over block-fading channels where the upper bounds on the ML decoding error probability are compared with computer simulations of the Log-MAP iterative decoding algorithm. This figure is reproduced from [122] with permission.

Fig. 4.9 illustrates the average block error probability of an ensemble of uniformly interleaved turbo codes whose transmission takes place over a block-fading Gaussian channel. The turbo encoder consists of two recursive systematic convolutional codes connected by a uniform interleaver of length 378 bits. The components of the turbo code have a common generator $G_1(D) = G_2(D) = [1, \frac{1+D+D^3}{1+D^2+D^3}, \frac{1+D+D^2+D^3}{1+D^2+D^3}]$, and the overall code rate is one-fifth bits per channel use. A coded frame is divided into an equal number of sub frames (1, 2, 3 or 4 in the figure), and the fading during each sub-frame is assumed to stay constant and be i.i.d. with a Rayleigh distribution; the noise added to the channel is assumed to be an additive Gaussian noise. The bound depicted in this figure relies on combining the SFB with the union bound (see [122]).

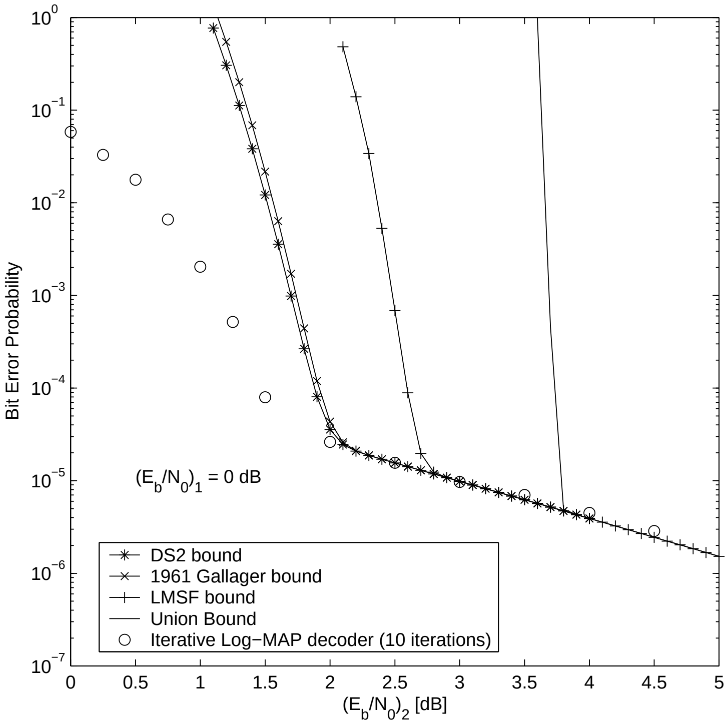

Fig. 4.10 Performance bounds for the bit error probability under ML decoding versus computer simulation results of iterative Log-MAP decoding (with 10 iterations). We refer here to the ensemble of uniformly interleaved turbo codes shown in Fig. 4.2 (see p. 137) where the transmission takes place over two (independent) parallel binary-input AWGN channels. Each bit is equally likely to be assigned to one of these channels, and the energy per bit to spectral noise density of the first channel is set to $\left(\frac{E_b}{N_0}\right)_1 = 0$ dB. The compared upper bounds on the bit error probability are the generalizations of the DS2 and 1961 Gallager bounds with their optimized tilting measures [166], the LMSF bound from [122] (which relies on a proper combination of the Shulman-Feder bound (SFB) and the union bound), and the union bound.

Tightened Gallager bounds on the performance of binary linear codes were recently derived for the case where the transmission takes place over independent parallel MBIOS channels [166, 165]. These bounds were exemplified for turbo-like codes with a special emphasis on accumulate-repeat-accumulate codes [1]. The first approach studied in [166, 165] is a generalization of the DS2 bound for parallel channels, and an optimization of the probability tilting measures in this bound with respect to each of these channels. The other approach is the optimization of the related tilting measures in the generalized 1961 Gallager bound where, except for the optimizations via calculus of variations, this bound was originally derived in [122]. The numerical results show a remarkable improvement over the special instances of the Gallager bounds studied in [122]. These bounds are exemplified for codes of finite length, and inner bounds on the asymptotic attainable channel regions are also computed. However, in some cases (see Fig. 4.10), the upper bounds which are valid under optimal ML decoding happen to be a bit pessimistic as compared to computer simulations of sub-optimal iterative decoding algorithms, thus indicating that there is room for further improvement.

4.8 Summary and conclusions

In this section, we discuss numerous efficient bounds on the decoding error probability of specific codes and ensembles under ML decoding, and we demonstrate the underlying connections that exist between them. In addressing the Gallager bounding techniques and their variations, we focus here on the Duman and Salehi variation, which originates from the classical 1965 Gallager bound. The considered upper bounds on the block and bit error probabilities under ML decoding rely on the distance spectra and the input-output weight distributions of the codes, respectively, which are in general calculable. By generalizing the second version of the Duman and Salehi (DS2) bound which was originally derived for the binary-input AWGN channel [60], we demonstrate its remarkable generality and show that it provides the natural bridge between the 1961 and 1965 Gallager bounds (see Section 4.3). It is applicable for both random and specific codes, as well as for either bit or block error analysis. Some observations and interconnections between the Gallager and Duman and Salehi bounds are presented in Section 4.3, which partially rely on insightful observations made by Divsalar [50]. In particular, it is demonstrated in Section 4.3 that the 1965 Gallager random coding bound can be re-derived from the DS2 bound. The geometric interpretations of these Gallager-type bounds are introduced in Section 4.3, reflecting the non-uniqueness of their associated probability tilting measures. In Section 4.4, many reported upper bounds on the ML decoding error probability (or their Chernoff versions) are shown to be special cases of the DS2 bound. This is done by choosing an appropriate probability tilting measure and calculating the resulting bound with the specific chosen measure. This framework also facilitates to generalize the Shulman-Feder bound (SFB) in [187] (see [183, Appendix A]).

The proposed approach can be generalized to geometrically uniform non-binary codes, finite-state channels, bit-interleaved coded modulation systems [35], and it can be also used for the derivation of upper bounds on the conditional decoding error probability (so as to account for a possible partitioning of the original code to subcodes). The DS2 bound is generalized in Section 4.5 to the mismatched decoding regime, where the decoder performs ML decoding with respect to a mismatched metric. The generalized bound is applicable to the analysis of the error performance of deterministic codes and ensembles. We address in particular the random coding version, which matches the 1965 Gallager random coding setting, and reproduces in Section 4.5 some known results (see [85], [106], [132]), hinging on an alternative approach which appropriately limits the considered ensemble of codes. Implications and applications of these observations are pointed out in Section 4.7, which include the fully interleaved fading channel with either matched or mismatched decoding.

The generalization of the DS2 upper bound for the analysis of the bit error probability yields the replacement of the distance spectrum $\{S_d\}_{d=1}^{N}$ (which appears in the upper bound on the block error probability) by the sequence $\{S'_d\}_{d=1}^{N}$ where: $S'_d = \sum_{w=1}^{NR} \left\{ \left(\frac{w}{NR}\right) A_{w,d} \right\}$, and $A_{w,d}$ designates the number of codewords in a systematic block code with an information Hamming weight w and a total Hamming weight d (where $0 \leq w \leq NR$ and $0 \leq d \leq N$). The derivation of the generalized DS2 bound on the bit error probability is detailed in [172], and it yields a considerable improvement in the tightness of the bound, as compared to [62, 60]. The computational complexity which is involved in calculating the DS2 bound is moderate.

The interconnections between many reported upper bounds are depicted in Fig. 4.11, where it is reflected that the DS2 bound yields many reported upper bounds as special cases.

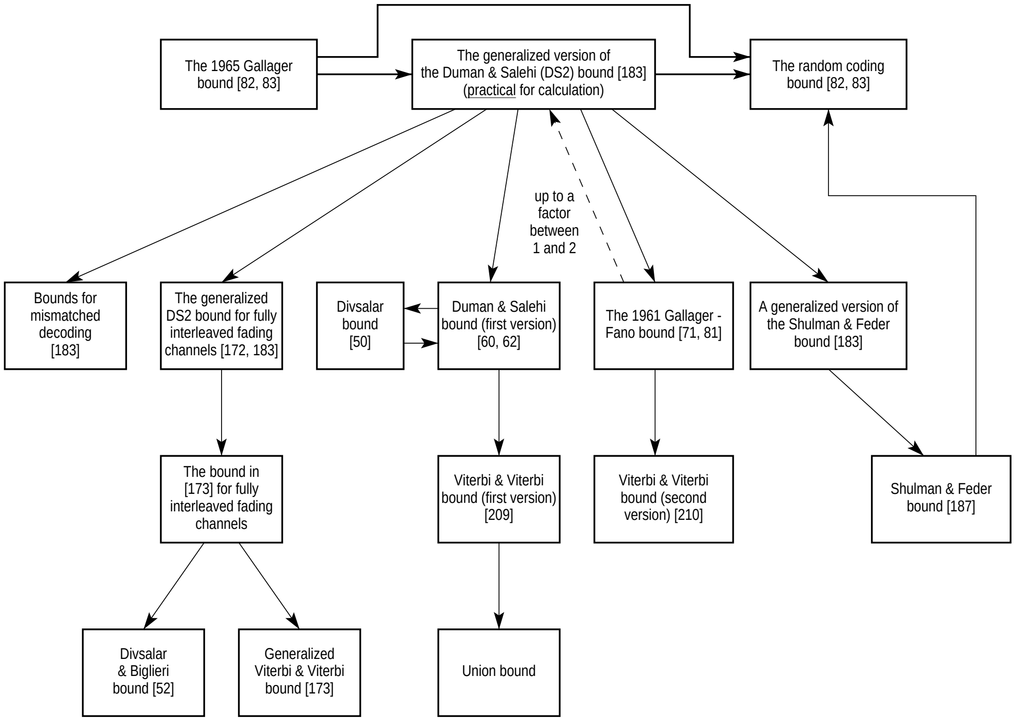
The 1965 Gallager bound [82, 83]
The generalized version of the Duman & Salehi (DS2) bound [183] (practical for calculation)
The random coding bound [82, 83]
up to a factor between 1 and 2
Bounds for mismatched decoding [183]
The generalized DS2 bound for fully interleaved fading channels [172, 183]
Divsalar bound [50]
Duman & Salehi bound (first version) [60, 62]
The 1961 Gallager - Fano bound [71, 81]
A generalized version of the Shulman & Feder bound [183]
The bound in [173] for fully interleaved fading channels
Viterbi & Viterbi bound (first version) [209]
Viterbi & Viterbi bound (second version) [210]
Shulman & Feder bound [187]
Divsalar & Biglieri bound [52]
Generalized Viterbi & Viterbi bound [173]
Union bound

Fig. 4.11 A diagram which shows interconnections among various upper bounds on the ML decoding error probability.

The tangential-sphere bound (TSB) is one of the tightest upper bounds for block codes which are transmitted over the binary-input AWGN channel and ML decoded (see Fig. 4.3, the discussion in Section 3.2, and [168, 170, 169, 164]). However, in the random coding setting, the TSB and some of its improved versions fail to reproduce the random coding error exponent (see Section 4.4.2.6 and [152, 199]), while the DS2 bound does. In fact, even the SFB which is a special case of the latter bound (see Section 4.4.1) achieves capacity for the ensemble of fully random block codes. This substantiates the claim that there is no uniformly best bound, providing the incentive for the generalization of the SFB (see [183, Appendix A] and some tightened versions of the SFB [200, 201]). However, we note that the loosened version of the TSB [50] (which involves the Chernoff inequality) maintains the asymptotic (i.e., for infinite block length) exponential tightness of the TSB of Poltyrev [152], and this loosened version is a special instance of the DS2 bound (see Section 4.4.2).

5

Sphere-Packing Bounds on the Decoding Error Probability: Classical and Recent Results

Overview: This section reviews the concepts used in the derivation of sphere-packing lower bounds on the block error probability. We review both the 1959 sphere-packing bound derived by Shannon for the Gaussian channel and the 1967 sphere-packing bound derived by Shannon, Gallager and Berlekamp for discrete memoryless channels. These ideas serve for presenting recent improvements on the tightness of the 1967 sphere-packing bound, as suggested by Valembois and Fossorier. These modifications provide an improved version of the classical 1967 sphere-packing bound whose tightness is enhanced especially for codes of short to moderate block lengths. These modifications also extend the application of the 1967 sphere-packing bound to memoryless continuous-output channels.

5.1 Introduction

The 1967 sphere-packing (SP67) bound, derived by Shannon, Gallager and Berlekamp [184], provides a lower bound on the decoding error probability of block codes as a function of their block length and their code rate, and it applies to arbitrary discrete memoryless channels (DMCs). Like the random coding bound of Gallager [82], the sphere-packing bound decays to zero exponentially with the block length for all rates below the channel capacity. Further, the error exponent of the SP67 bound is a convex function of the rate which is known to be tight at the portion of the rate region between the critical rate (R_c) and the channel capacity; for this important rate region, the error exponents of the random coding and SP67 bounds coincide [184, Part 1]. The 1959 sphere-packing (SP59) bound, derived by Shannon [185], provides a lower bound on the block error probability when the transmission takes place over an AWGN channel. Its derivation relies on first showing that the error probability of any code whose codewords lie on a sphere must be greater than the error probability of a code of the same length and rate whose codewords are uniformly distributed over that sphere.

The SP67 bound happens to be loose for codes of moderate block lengths. This is due to the original focus in [184] on asymptotic analysis. In their paper [204], Valembois and Fossorier revisited the SP67 bound in order to improve its tightness for codes of short to moderate block lengths, and also to extend its spectrum of applications to continuous-output channels (e.g., the AWGN channel which is the communication channel model for the SP59 bound of Shannon [185]). The motivation for the study in [204] was strengthened due to the outstanding performance of codes defined on graphs even for moderate block lengths. The remarkable improvement in the tightness of the SP67 bound was exemplified in [204] for the case of BPSK signaling over the AWGN channel, and it was shown that the improved version of the SP67 bound provides an interesting alternative to the SP59 bound [185].

This section is structured as follows: Section 5.2 presents the SP59 bound of Shannon [185] which provides a lower bound on the ML decoding error probability for the Gaussian channel in terms of the code rate and the block length of the code; this section also addresses some tips from [204] which are provided for simplifying the computation of the SP59 bound. Section 5.3 presents the classical SP67 bound which was originally derived by Shannon, Gallager and Berlekamp for a general discrete memoryless channel (DMC) [184], and explains the concepts and main ideas which were used for the derivation of the SP67 bound. Section 5.4 provides an improved version of the SP67 bound, as suggested by Valembois and Fossorier in [204], which makes the bound tighter for codes of moderate block length. The new bound also extends the validity of this bounding technique to memoryless continuous-output channels (e.g., the Gaussian channel, for which the SP59 bound applies). For the Gaussian channel, some numerical results from [204], comparing between the SP59 bound and the improved SP67 bound, are presented. Concluding comments on sphere-packing bounds are finally given in Section 5.5.

5.2 The 1959 Shannon lower bound for the AWGN channel

The SP59 bound of Shannon [185] provides a lower bound on the decoding error probability of an arbitrary block code whose transmission takes place over an AWGN channel. Consider a block code $\mathcal{C}$ of length N, and assume that the rate of the code is R bits per channel use per dimension. It is assumed that all the codewords are mapped to signals of the same energy, but the bound does not take into account the particular modulation which is used. Since all the signals are represented by points on a sphere centered at the origin, then every Voronoi cell is a polyhedric cone which is limited by at most $2^{NR} - 1$ hyper planes intersecting at the origin.

Definition 5.1. (Solid angle of a cone) The solid angle of a cone is defined as the area of a sphere of unit radius (in the same Euclidean space) cut out by the cone.

Since the 2^{NR} Voronoi cells partition the Euclidean space $\mathbb{R}^N$, the sum of their solid angles is equal to the solid angle of the whole space, i.e., it is equal to the total area of a sphere of unit radius in the N-dimensional space.

The main idea used for the derivation of the SP59 bound is that the error probability of an arbitrary code under ML decoding is lower bounded by the decoding error probability which corresponds to the case where the Voronoi regions are circular cones of equal solid angles; their common value is equal to a fraction 2^{-NR} of the solid angle of $\mathbb{R}^N$. This follows as a corollary of the two propositions (see [185])

- Among the cones of a given solid angle, the circular cone provides the lowest decoding error probability under ML decoding.
- In order to minimize the average decoding error probability, it is best to share the total solid angle equally between the 2^{NR} Voronoi cells.

Lemma 5.2. (Solid angle of a circular cone) The solid angle of a circular cone of half angle θ in $\mathbb{R}^N$ is given by

$$\Omega_N(\theta) = \frac{2\pi^{\frac{N-1}{2}}}{\Gamma(\frac{N-1}{2})} \int_0^\theta (\sin\phi)^{N-2}\, d\phi\,. \tag{5.1}$$

In particular, the solid angle of the whole space is

$$\Omega_N(\pi) = \frac{2\pi^{\frac{N}{2}}}{\Gamma(\frac{N}{2})}\,. \tag{5.2}$$

Theorem 5.3. (The SP59 bound) Let

$$f_n(x) = \frac{1}{2^{(n-1)/2}\Gamma(\frac{n+1}{2})} \int_0^\infty z^{n-1} \exp(-\frac{z^2}{2} + zx)\, dz\,, \quad \forall\, x \in \mathbb{R},\, n \in \mathbb{N} \tag{5.3}$$

and

$$\begin{aligned} P_{\text{SPB}}(N,\theta,A) = {} & \frac{(N-1)e^{-\frac{NA^2}{2}}}{\sqrt{2\pi}} \int_\theta^{\frac{\pi}{2}} (\sin\phi)^{N-2}\, f_N(\sqrt{N}A\cos\phi)\, d\phi \\ & + Q(\sqrt{N}A). \end{aligned} \tag{5.4}$$

Then, the ML decoding error probability of any code of block length N and rate R satisfies the following lower bound:

$$P_{\text{e}}(\text{ML}) > P_{\text{SPB}}(N,\theta,A)\,, \quad A \triangleq \sqrt{\frac{2E_{\text{s}}}{N_0}} \tag{5.5}$$

for all $\theta \in [0,\pi]$ so that $2^{-NR} \leq \frac{\Omega_N(\theta)}{\Omega_N(\pi)}$.

The signal point which is represented by the transmitted signal lies on a sphere of radius $\sqrt{NE_\mathrm{s}}$ in $\mathbb{R}^N$. The value $P_{\mathrm{SPB}}(N,\theta,A)$ in the RHS of (5.5) designates the probability that the additive noise vector moves this signal point outside a cone of half angle θ whose main axis passes through the origin and the signal point. Hence, this function is monotonically decreasing in θ. The tightest lower bound on the decoding error probability is therefore achieved for $\theta_1(N,R)$ which satisfies

$$\frac{\Omega_N\big(\theta_1(N,R)\big)}{\Omega_N(\pi)} = 2^{-NR}.$$

The calculation of $\theta_1(N,R)$ may become quite tedious for large values of N. In order to simplify the calculation of the SP59 bound, [185] provides asymptotically tight upper and lower bounds on the ratio $\frac{\Omega_N(\theta)}{\Omega_N(\pi)}$. These bounds are provided in the following lemma:

Lemma 5.4. (Bounds on the solid angle) The solid angle of a circular cone of half angle θ in the Euclidean space $\mathbb{R}^N$ satisfies the inequality

$$\frac{\Gamma(\frac{N}{2})(\sin\theta)^{N-1}}{2\Gamma(\frac{N+1}{2})\sqrt{\pi}\cos\theta}\left(1-\frac{(\tan\theta)^2}{N}\right) \le \frac{\Omega_N(\theta)}{\Omega_N(\pi)} \le \frac{\Gamma(\frac{N}{2})(\sin\theta)^{N-1}}{2\Gamma(\frac{N+1}{2})\sqrt{\pi}\cos\theta}\,. \quad (5.6)$$

Corollary 5.5. (SP59 bound (cont.)) If θ^* satisfies the equation

$$\frac{\Gamma(\frac{N}{2})(\sin\theta^*)^{N-1}}{2\Gamma(\frac{N+1}{2})\sqrt{\pi}\cos\theta^*}\left(1-\frac{(\tan\theta^*)^2}{N}\right) = 2^{-NR} \quad (5.7)$$

then $\frac{\Omega_N(\theta^*)}{\Omega_N(\pi)} \ge 2^{-NR}$, and thus

$$P_\mathrm{e}(\mathrm{ML}) > P_{\mathrm{SPB}}(N,\theta^*,A) \quad (5.8)$$

The use of θ^* instead of the optimal value $\theta_1(N,R)$ causes some loss in the tightness of the SP59 bound. However, due to the asymptotic tightness of the bounds on $\frac{\Omega_N(\theta)}{\Omega_N(\pi)}$, the loss in the tightness of the bound in Corollary 5.5 vanishes asymptotically as we let N tend to infinity. In [204], it was numerically observed that the loss is marginal even for small values of the dimension (NR). For example, it was observed that the loss is smaller then 0.01 dB whenever the dimension of the code is greater than 20, and becomes even smaller then 0.001 dB when the dimension of the code exceeds 60.

In [204, Section 2], the SP59 bound was reviewed, and an algorithm which simplifies the calculation of the bound is given in the following theorem:

Theorem 5.6. (Recursive equations for simplifying the calculation of the SP59 bound) [204, Theorem 3]: The set of functions $\{f_n\}$ introduced in (5.3) can be rewritten in the alternative form

$$f_n(x) = P_n(x) + Q_n(x) \exp(\frac{x^2}{2}) \int_{-\infty}^{x} \exp(-\frac{t^2}{2})\, dt\ , \quad x \in \mathbb{R},\ n \in \mathbb{N} \tag{5.9}$$

where P_n and Q_n are two polynomials which can be determined by the same recursive equation for all $n \geq 5$:

$$\begin{aligned} P_n(x) &= \frac{2n-5+x^2}{n-1}\, P_{n-2}(x) - \frac{n-4}{n-1}\, P_{n-4}(x)\ , \\ Q_n(x) &= \frac{2n-5+x^2}{n-1}\, Q_{n-2}(x) - \frac{n-4}{n-1}\, Q_{n-4}(x) \end{aligned} \tag{5.10}$$

with the initial conditions

$$\begin{aligned} &P_1(x) = 0, \quad Q_1(x) = 1 \\ &P_2(x) = \sqrt{\frac{2}{\pi}}, \quad Q_2(x) = \sqrt{\frac{2}{\pi}}\, x \\ &P_3(x) = \frac{x}{2}, \quad Q_3(x) = \frac{1+x^2}{2} \\ &P_4(x) = \sqrt{\frac{2}{\pi}} \frac{2+x^2}{3}, \quad Q_4(x) = \sqrt{\frac{2}{\pi}} \frac{3x+x^3}{3}. \end{aligned} \tag{5.11}$$

This theorem is proved in [204, Appendix A]. Note that the algorithm in this theorem can be applied to the calculation of the SP59 bound for values of N not exceeding 1,000 (due to numerical problems of overflows and underflows which are related to the calculation of the recursive equations (5.10) for larger values of N). For numerical accuracy purposes, it was suggested in [204] to use the two recursive equations in (5.10) with their initial values in (5.11) in order to precompute the coefficients of the two polynomials P_N and Q_N (i.e., by simply performing convolution operations); note that the function

$$y(x) = \exp(\frac{x^2}{2}) \int_{-\infty}^{x} \frac{\exp(-\frac{t^2}{2})}{\sqrt{2\pi}} \, dt$$

appearing as a multiplicative function at the second term of the RHS in (5.9) can be easily calculated using standard programming languages; e.g., with the MATLAB software, the function y can be expressed in the form

$$y(x) = \text{erfcx}(-\frac{x}{\sqrt{2}}) \, , \quad x \in \mathbb{R}$$

and for $x \ll -1$, it is well approximated by $y(x) \approx -\frac{1}{\sqrt{2\pi}\, x}$.

For large block lengths, [185] presents some asymptotic formulas which give a very accurate estimation of the bound for large enough block lengths. These approximations allow the calculation to be made in the logarithmic domain which virtually eliminates the possibility of floating point errors.

Theorem 5.7. [185]: Defining

$$G(\theta) \triangleq \frac{A\cos\theta + \sqrt{A^2\cos^2\theta + 4}}{2}$$
$$E_L(\theta) \triangleq \frac{A^2 - AG(\theta)\cos\theta - 2\ln(G(\theta)\sin\theta)}{2}$$

then

$$P_{\text{SPB}}(N,\theta,A) \geq \frac{\sqrt{N-1}}{6N(A+1)} e^{\frac{-(A+1)^2+3}{2}} e^{-N\, E_L(\theta)}. \tag{5.12}$$

This lower bound is valid for any block length N. However, the ratio of the terms in the LHS and RHS of (5.12) stays bounded away from one for all values of N. A better approximation of $P_{\text{SPB}}(N,\theta,A)$ is given by the next theorem, but without a determined inequality. As a consequence, the following approximation is not a proven theoretical lower bound on the error probability. For $N > 1000$, however, its numerical values become almost identical to those of the exact bound, thus giving a useful approximation to this lower bound (see also [57]).

Theorem 5.8. [185]: Using the notation of Theorem 5.7, if $\theta > \cot^{-1}(A)$, then

$$P_{\mathrm{SPB}}(N,\theta,A) \approx \frac{\alpha(\theta)e^{-NE_L(\theta)}}{\sqrt{N}}$$

where

$$\alpha(\theta) \triangleq \left(\sqrt{\pi\bigl(1+G(\theta)^2\bigr)}\sin\theta\bigl(AG(\theta)\sin^2\theta - \cos\theta\bigr)\right)^{-1}.$$

Fig. 5.1 which was reproduced from [204] refers to the $(128,64,22)$ extended BCH codes; it compares the SP59 bound with the actual performance achieved by a near-ML decoding algorithm when transmission takes place over an AWGN channel and the codewords are BPSK modulated (for further details about this sub-optimal decoding algorithm, the reader is referred to [203, Section 6B]). This figure shows a small

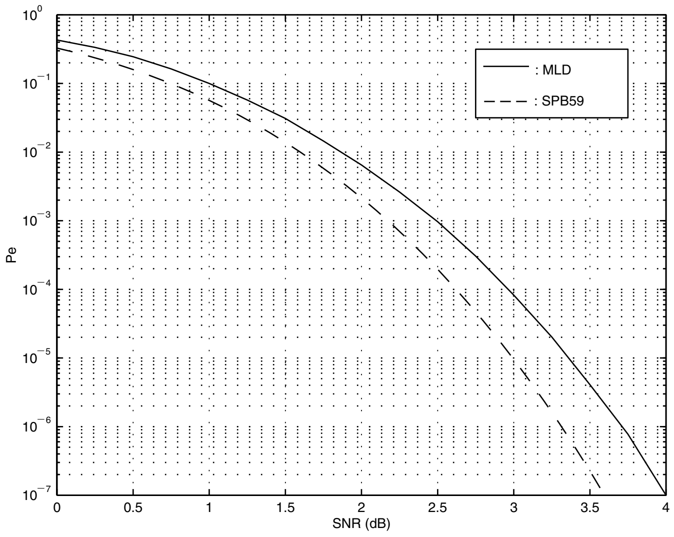

Fig. 5.1 A comparison between the 1959 sphere-packing (SP59) bound of Shannon [185] for $(128,64)$ block codes and the ML performance of the extended $(128,64,22)$ BCH code over the AWGN channel with BPSK modulation. The performance of the extended BCH code relies on the BMA algorithm from [203] which is a near-ML decoding algorithm with reduced complexity. This figure is reproduced (with permission) from [204]. For further details, the reader is referred to [203, Section 6B].

gap of only 0.34 dB for a target block error probability of $P_{\mathrm{t}} = 10^{-4}$ between the SP59 bound and the performance of the considered rate-$\frac{1}{2}$ code of block length 128 bits.

A comparison of the performance of turbo codes under iterative sum-product decoding algorithm with the SP59 lower bound shows (see [57, 115, 125, 181]) the existence of good turbo codes of rate between $\frac{1}{6}$ and $\frac{1}{2}$ which are about 0.7 dB away of the SP59 bound for a target error probability of 10^{-4} and moderate block lengths.

5.3 The 1967 sphere-packing bound

We outline here the concepts and main ideas which serve for the derivation of the SP67 bound, as originally introduced by Shannon, Gallager and Berlekamp [184]. The SP67 bound applies to all DMCs, and it possesses the very pleasing feature that its error exponent is exact for all rates between the critical rate and channel capacity [184, Part I]; this nice property also holds in the limit where the code rate tends to zero [184, Part II]. The analysis in [184] was primarily focused on the exponential behavior of the SP67 bound, and the aim of the authors was to simplify the derivation of the bound as long as no penalty is incurred in its exponential behavior.

Apart of outlining the classical derivation of the SP67 bound, the discussion in this section also serves as a preparatory step towards the consideration of possible refinements of the SP67 bound in order to make it attractive for moderate block lengths. Hence, in the next section, we rely on the proof concepts introduced in this section, and use them to explain the improvements on the SP67 bound, as suggested in [204]. By doing so, Valembois and Fossorier have managed to adapt their new sphere-packing bound so that it also applies to memoryless channels with continuous output (and not only to DMCs); their new bound also competes well with the 1959 sphere-packing bound of Shannon [185]. The study in [204], extending the applicability of the sphere-packing bounding technique to codes of moderate block lengths, was highly motivated by the outstanding performance of turbo-like codes which obtain reliable communication at rates close to capacity with tolerable latency and complexity (even under sub-optimal iterative decoding algorithms). To conclude, we start our discussion by outlining the concepts of the proofs used for the classical derivation of the SP67 bound [184], and then move in the next section to an improved version of this bound for moderate block lengths and for memoryless channels with continuous outputs. We consider this study to be of primary significance in the evaluation of how close to optimal are turbo-like codes together with their (sub-optimal) iterative decoding algorithms (in this respect, the reader is also referred to [57], [115] and [125]).

For the derivation of the classical SP67 bound [184], let $\mathbf{x}_1, \ldots, \mathbf{x}_M$ be a set of M codewords of length N, and assume that their transmission takes place over a DMC. Also let us assume a list decoder where for each received sequence $\mathbf{y}$, the decoder outputs a list of at most L integers from 1 to M; these integers correspond to the indices of the codewords (or the corresponding messages before encoding). When the codeword $\mathbf{x}_m$ is transmitted over the channel, an error is declared if the index m does not appear in the list. In [184], the authors derive a lower bound on the decoding error probability of a code with M codewords of block length N, assuming an arbitrary list decoder whose size is limited to L. The particular case where $L = 1$ clearly provides a lower bound on the performance under optimal ML decoding (since the list size is limited to one for each received sequence at the channel output). Referring to a list decoder of size at most L, the code rate (in nats per symbol use) is given by $R = \frac{\ln\left(\frac{M}{L}\right)}{N}$; the motivation for this definition is that the possible number of messages is reduced from M to L, thus $\ln\left(\frac{M}{L}\right)$ nats of information are gained. For the case where $L = 1$, this definition particularizes to the standard definition of code rate. In the other extreme case where $L = M$, all the codewords are in the list for any output sequence $\mathbf{y}$, so there is no need for a communication channel and the transmission rate is therefore zero.

Let $\mathcal{Y}_m$ be the set of output sequences $\mathbf{y}$ for which the message index m is on the decoding list. The probability of error under list decoding when the message whose index is m is sent over the DMC, is given by

$$P_{\mathrm{e},m} = \sum_{y \in \mathcal{Y}_m^{\mathrm{c}}} \Pr\left(\mathbf{y}|\mathbf{x}_m\right) \tag{5.13}$$

where the superscript c designates the complementary set. Let $P_{\mathrm{e,max}}$ designate the maximum over m of $P_{\mathrm{e},m}$ for the code and list decoding under consideration. Assuming that the codewords are equally likely to be transmitted, the average decoding error probability is given by

$$P_{\mathrm{e}} = \frac{1}{M} \sum_{m=1}^{M} P_{\mathrm{e},m}. \tag{5.14}$$

In [184], the derivation of the SP67 bound was divided into three steps where in the first step, the authors derive upper and lower bounds on the decoding error probability for a block code with simply two codewords; as a second step, a lower bound on the decoding error probability is derived for fixed composition codes, and finally, the derivation of the SP67 bound is completed for an arbitrary block code under list decoding. In the following, we outline the derivation of the SP67 bound according to these steps.

Upper and lower bounds for a code with two codewords

Let $P_m(\mathbf{y})$ (where $m = 1, 2$) be the probability of receiving the sequence $\mathbf{y}$ when message no. m is transmitted. Lets assume ML decoding, and let $\mathcal{Y}_m$ be the set of sequences decoded into the message index m. Summing (5.13) over m gives the equality

$$P_{\mathrm{e},1} + P_{\mathrm{e},2} = \sum_{\mathbf{y}} \min_{m=1,2} P_m(\mathbf{y}).$$

In order to obtain upper and lower bounds on the two conditional error probabilities $P_{\mathrm{e},1}$ and $P_{\mathrm{e},2}$ and study the tradeoff between these two error probabilities, the authors rely on a simple and useful inequality which is later shown to be asymptotically tight in the limit where N tends to infinity. The inequality states that for all the received vectors $\mathbf{y}$

$$\min_{m=1,2} P_m(\mathbf{y}) \le P_1(\mathbf{y})^{1-s} P_2(\mathbf{y})^s \le \max_{m=1,2} P_m(\mathbf{y}) \, , \quad 0 < s < 1$$

hence, the combination of the last two equations gives

$$P_{\mathrm{e},1} + P_{\mathrm{e},2} \le e^{\mu(s)}, \quad 0 < s < 1. \tag{5.15}$$

where

$$\mu(s) \triangleq \ln\Big(\sum_{\mathbf{y}} P_1(\mathbf{y})^{1-s} P_2(\mathbf{y})^s\Big). \tag{5.16}$$

By extending the definition of the function μ to the interval [0,1]:

$$\mu(0) \triangleq \lim_{s\to 0^+} \mu(s), \quad \mu(1) \triangleq \lim_{s\to 1^-} \mu(s)$$

so as to cover the endpoints $s = 0$ and $s = 1$, one obtains from (5.15) the inequality

$$P_{\mathrm{e},1} + P_{\mathrm{e},2} \le \min_{0\le s\le 1} e^{\mu(s)}. \tag{5.17}$$

Let the codewords be denoted by $\mathbf{x}_m = (k_{m,1}, \dots, k_{m,N})$ where $m = 1, 2$, and let the received sequence be $\mathbf{y} = (j_1, \dots, j_N)$. Since the communication channel is memoryless, then $P_m(\mathbf{y}) = \prod_{n=1}^N \Pr(j_n|k_{m,n})$, and the function μ is expressible in the form

$$\mu(s) = \sum_{n=1}^N \mu_n(s), \quad \mu_n(s) \triangleq \ln\Big(\sum_{j=1}^J \Pr(j|k_{1,n})^{1-s} \Pr(j|k_{2,n})^s\Big). \tag{5.18}$$

The next step in [184] is the derivation of upper and lower bounds on $P_{\mathrm{e},1}$ and $P_{\mathrm{e},2}$, where the authors allow themselves the flexibility of making $P_{\mathrm{e},1}$ very much larger than $P_{\mathrm{e},2}$ or vice versa. The authors prove the following theorem:

Theorem 5.9. (Upper and lower bounds for the pairwise error probability) [184, Theorem 5]: Let $P_1(\mathbf{y})$ and $P_2(\mathbf{y})$ be two probability assignments on a discrete set of sequences, let $\mathcal{Y}_1$ and $\mathcal{Y}_2$ be disjoint decision regions for these sequences, let $P_{\mathrm{e},1}$ and $P_{\mathrm{e},2}$ be given by (5.13), and assume that $P_1(\mathbf{y})P_2(\mathbf{y}) \neq 0$ for at least one sequence $\mathbf{y}$. Then, for all $s \in (0, 1)$, either

$$P_{\mathrm{e},1} > \frac{1}{4} \exp\Big(\mu(s) - s\mu'(s) - s\sqrt{2\mu''(s)}\Big) \tag{5.19}$$

or

$$P_{\mathrm{e},2} > \frac{1}{4} \exp\Big(\mu(s) + (1-s)\mu'(s) - (1-s)\sqrt{2\mu''(s)}\Big) \tag{5.20}$$

where the function μ is introduced in (5.16). Furthermore, for an appropriate choice of $\mathcal{Y}_1$ and $\mathcal{Y}_2$

$$P_{\mathrm{e},1} \le \exp\Big(\mu(s) - s\mu'(s)\Big) \tag{5.21}$$

and

$$P_{\mathrm{e},2} \le \exp\Big(\mu(s) + (1-s)\mu'(s)\Big). \tag{5.22}$$

Finally, the function μ is non-positive and convex over the interval $(0, 1)$. The convexity of μ is strict unless $\frac{P_1(\mathbf{y})}{P_2(\mathbf{y})}$ is constant over all the sequences $\mathbf{y}$ for which $P_1(\mathbf{y})P_2(\mathbf{y}) \neq 0$. Moreover, the function μ is strictly negative over the interval $(0, 1)$ unless $P_1(\mathbf{y}) = P_2(\mathbf{y})$ for all $\mathbf{y}$.

Concept of the proof: Taking the first and second derivatives of the function μ gives

$$\mu'(s) = \sum_{\mathbf{y}} Q_s(\mathbf{y}) D(\mathbf{y}), \quad \mu''(s) = \left(\sum_{\mathbf{y}} Q_s(\mathbf{y}) D(\mathbf{y})^2\right) - \mu'(s)^2 \tag{5.23}$$

where Q_s is defined to be the probability measure

$$Q_s(\mathbf{y}) = \frac{P_1(\mathbf{y})^{1-s} P_2(\mathbf{y})^s}{\sum_{\mathbf{y}'} P_1(\mathbf{y}')^{1-s} P_2(\mathbf{y}')^s}$$

and $D(\mathbf{y}) \triangleq \ln \frac{P_2(\mathbf{y})}{P_1(\mathbf{y})}$ designates the log-likelihood ratio. If we consider $D(\mathbf{y})$ to be a RV with probability assignment $Q_s(\mathbf{y})$, then it follows from (5.23) that the first and second derivatives of μ are equal to the expectation and variance of the RV $D(\mathbf{y})$, respectively, i.e.,

$$\mu'(s) = \mathbb{E}_{Q_s}\big\{D(\mathbf{y})\big\}, \quad \mu''(s) = \mathrm{Var}_{Q_s}\big\{D(\mathbf{y})\big\}.$$

Hence, $\mu''(s) \ge 0$, so the function μ is convex (and it is strictly convex if and only if $D(\mathbf{y})$ is not a constant for all the sequences $\mathbf{y}$ for which $P_1(\mathbf{y})P_2(\mathbf{y}) \neq 0$). It is direct to show that the probability assignments P_1 and P_2 are expressible in the form

$$P_1(\mathbf{y}) = \exp\big(\mu(s) - sD(\mathbf{y})\big)\, Q_s(\mathbf{y}) \tag{5.24}$$

$$P_2(\mathbf{y}) = \exp\big(\mu(s) + (1-s)D(\mathbf{y})\big)\, Q_s(\mathbf{y}). \tag{5.25}$$

For the derivation of the lower bounds, the authors define in [184] the parameterized set

$$\begin{aligned}\widetilde{\mathcal{Y}}_s &\triangleq \Big\{\mathbf{y} : |D(\mathbf{y}) - \mu'(s)| \le \sqrt{2\mu''(s)}\Big\}\\ &= \Big\{\mathbf{y} : |D(\mathbf{y}) - \mathbb{E}_{Q_s}\big\{D(\mathbf{y})\big\}|^2 \le 2\mathrm{Var}_{Q_s}\big\{D(\mathbf{y})\big\}\Big\}\end{aligned} \tag{5.26}$$

hence, from the Chebychev inequality

$$\sum_{\mathbf{y}\in\widetilde{\mathcal{Y}}_s} Q_s(\mathbf{y}) > \frac{1}{2}.$$

The idea of the derivation of the lower bounds on $P_{\mathrm{e},1}$ and $P_{\mathrm{e},2}$ is to take into account only the error events where the received sequence $\mathbf{y}$ falls within the set $\widetilde{\mathcal{Y}}_s$ defined in (5.26). From (5.24) and (5.26)

$$\begin{aligned} P_{\mathrm{e},1} &= \sum_{\mathbf{y}\in\mathcal{Y}_1^{\mathrm{c}}} P_1(\mathbf{y}) \\ &\geq \sum_{\mathbf{y}\in\mathcal{Y}_1^{\mathrm{c}}\cap\widetilde{\mathcal{Y}}_s} P_1(\mathbf{y}) \\ &\geq \exp\Bigl(\mu(s) - s\mu'(s) - s\sqrt{2\mu''(s)}\Bigr) \sum_{\mathbf{y}\in\mathcal{Y}_\mathbf{1}^{\mathrm{c}}\cap\widetilde{\mathcal{Y}}_\mathbf{s}} Q_s(\mathbf{y}) \end{aligned} \qquad (5.27)$$

and similarly

$$P_{\mathrm{e},2} \geq \exp\Bigl(\mu(s) + (1-s)\mu'(s) - (1-s)\sqrt{2\mu''(s)}\Bigr) \sum_{\mathbf{y}\in\mathcal{Y}_\mathbf{2}^{\mathrm{c}}\cap\widetilde{\mathcal{Y}}_\mathbf{s}} Q_s(\mathbf{y}). \qquad (5.28)$$

Finally, since the sets $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are assumed to be disjoint, then

$$\sum_{\mathbf{y}\in\mathcal{Y}_\mathbf{1}^{\mathrm{c}}\cap\widetilde{\mathcal{Y}}_\mathbf{s}} Q_s(\mathbf{y}) + \sum_{\mathbf{y}\in\mathcal{Y}_\mathbf{2}^{\mathrm{c}}\cap\widetilde{\mathcal{Y}}_\mathbf{s}} Q_s(\mathbf{y}) = \sum_{\mathbf{y}\in\widetilde{\mathcal{Y}}_s} Q_s(\mathbf{y}) > \frac{1}{2}$$

so at least one of the terms $\sum_{\mathbf{y}\in\mathcal{Y}_\mathbf{1}^{\mathrm{c}}\cap\widetilde{\mathcal{Y}}_\mathbf{s}} Q_s(\mathbf{y})$ and $\sum_{\mathbf{y}\in\mathcal{Y}_\mathbf{2}^{\mathrm{c}}\cap\widetilde{\mathcal{Y}}_\mathbf{s}} Q_s(\mathbf{y})$ should be larger than $\frac{1}{4}$, which finally leads to the validity of at least one of the lower bounds in (5.19) and (5.20).

For the derivation of the upper bounds in (5.21) and (5.22), let

$$\mathcal{Y}_1 \triangleq \bigl\{\mathbf{y} \,:\, D(\mathbf{y}) < \mu'(s)\bigr\}, \quad \mathcal{Y}_2 \triangleq \mathcal{Y}_1^{\mathrm{c}} \qquad (5.29)$$

so that the sets $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are disjoint. The derivation of the upper bounds in (5.21) and (5.22) follows from (5.24) and (5.25), and it relies on the definition of the disjoint sets $\mathcal{Y}_1$ and $\mathcal{Y}_2$ in (5.29). To obtain the upper bounds, one simply replaces the sums $\sum_{\mathbf{y}\in\mathcal{Y}_m} Q_s(\mathbf{y})$ by 1.

Concepts of proof for the derivation of the SP67 bound for fixed composition codes

Following the derivation of upper and lower bounds on the pairwise error probability (see Theorem 5.9), Shannon, Gallager and Berlekamp derived in [184] a lower bound on the decoding error probability for *fixed composition codes*. Let $\mathcal{C}$ be a block code of M codewords, each of length N, and assume that the transmission of this code takes place over a DMC whose probability law is given by $P(j|k)$ where $k \in \{1, \ldots, K\}$ and $j \in \{1, \ldots, J\}$ designate the input and output symbols, respectively. For the decoding of the code $\mathcal{C}$, an arbitrary list decoder is assumed where the size of the list is limited to L for any output sequence $\mathbf{y}$. By assumption, since the code $\mathcal{C}$ has a fixed composition, then each symbol $k \in \{1, \ldots, K\}$ appears in every one of the codewords $\mathbf{x}_1, \ldots, \mathbf{x}_M$ in $\mathcal{C}$ an equal number of times. Let us define the vector $\mathbf{q} \triangleq (q_1, \ldots, q_K)$ to be a K-length vector where q_k $(k = 1, \ldots, K)$ designates the fraction of appearances of the symbol k in each one of the codewords $\mathbf{x}_1, \ldots, \mathbf{x}_M$. Let $P_{\mathrm{e},m}$, as given in (5.13), be the error probability under list decoding given that the index of the transmitted message is m, and let $P_{\mathrm{e,max}}$ designate the maximum of $P_{\mathrm{e},m}$ over m (where $m = 1, \ldots, M$). For a given m, one can reduce the value of $P_{\mathrm{e},m}$ by enlarging the set $\mathcal{Y}_m$ of output sequences for which the index message m is on the list of the decoder. However, since the size of the list is limited to L for any output sequence, this will decrease the size of the set $\mathcal{Y}_{m'}$ for some $m' \neq m$, and hence, will increase the value of $P_{\mathrm{e},m'}$. In order to keep some control on the size of the set $\mathcal{Y}_m$ without specifically considering the other codewords, the authors in [184] define an arbitrary probability tilting measure which can be expressed in the product form

$$f_N(\mathbf{y}) = \prod_{n=1}^{N} f(j_n) \tag{5.30}$$

for any output sequence $\mathbf{y} = (j_1, \ldots, j_N)$, and define accordingly the *size* of the set $\mathcal{Y}_m$ as

$$F(\mathcal{Y}_m) \triangleq \sum_{\mathbf{y} \in \mathcal{Y}_m} f_N(\mathbf{y}). \tag{5.31}$$

Next, they rely on Theorem 5.9 in order to relate the conditional error probability $P_{\mathrm{e},m}$ and $F(\mathcal{Y}_m)$; to this end, let $\Pr(\mathbf{y}|\mathbf{x}_m)$ and $f_N(\mathbf{y})$ correspond to $P_1(\mathbf{y})$ and $P_2(\mathbf{y})$ in Theorem 5.9, respectively. In light of these substitutions, the function μ in (5.16) gets the form

$$\mu(s) \triangleq \ln\Bigl(\sum_{\mathbf{y}} \Pr(\mathbf{y}|\mathbf{x}_m)^{1-s} f_N(\mathbf{y})^s\Bigr) \tag{5.32}$$

and similarly to (5.18), since the tilting measure f_N is expressed in the product form (5.30) and the code is assumed to have a fixed composition, then the function μ in (5.32) is expressible by the sum

$$\mu(s) = N\sum_{k=1}^{K} q_k \mu_k(s,f), \quad \mu_k(s,f) \triangleq \ln\left(\sum_{j=1}^{J} P(j|k)^{1-s} f(j)^s\right). \tag{5.33}$$

By doing so, Theorem 5.9 provides a lower bound on either $P_{\mathrm{e},m}$ or $F(\mathcal{Y}_m)$ which is valid for every $m \in \{1,\dots,M\}$. Now, since the size of the list at the decoder is limited to L, then from (5.31)

$$\sum_{m=1}^{M} F(\mathcal{Y}_m) = \sum_{m=1}^{M} \sum_{\mathbf{y}\in\mathcal{Y}_\mathbf{m}} f_N(\mathbf{y}) \le \sum_{l=1}^{L}\sum_{\mathbf{y}} f_N(\mathbf{y}) = L \tag{5.34}$$

so there exists an integer $m \in \{1,\dots,M\}$ for which $F(\mathcal{Y}_m) \le \frac{L}{M}$. Moreover, for this value of m, one can simply replace a lower bound on $P_{\mathrm{e},m}$ with a lower bound on $P_{\mathrm{e,max}}$. Based on the additivity property of the function μ in (5.18), the term $\sqrt{\mu''(s)}$ in the exponents of the lower bounds in (5.19) and (5.20) is proportional to the square root of N while the other terms $\mu(s)$ and $\mu'(s)$ in these exponents are linearly proportional to N. Hence, since Shannon, Gallager and Berlekamp mainly focused in their paper [184] on the exponential behavior of the upper and lower bounds for large values of N, they chose for the simplicity of their analysis to obtain a loose but general upper bound on $\sqrt{\mu''(s)}$. The exponential behavior of the bounds remains unaffected in the asymptotic case where N tends to infinity. However, as we see in the next section, this bound is loose when considering codes of moderate block lengths. A loose but simple bound on the second derivative of the function μ was derived in [184, Appendix B], and it reads

$$s\sqrt{\mu_k''(s,f)} \le \ln\Bigl(\frac{e}{\sqrt{P_{\min}}}\Bigr) \tag{5.35}$$

where $P_{\min}$ designates the smallest non-zero transition probability of the DMC.

Following this approach and using the additivity property of the function μ in (5.33) and inequality (5.35), one obtains from Theorem 5.9 that if there exists a sequence $\mathbf{y}$ for which $\Pr(\mathbf{y}|\mathbf{x}_m) f_N(\mathbf{y}) \neq 0$, then either

$$P_{\mathrm{e,max}} > \exp\left\{N\left[\sum_{k=1}^{K} q_k\big(\mu_k(s,f) - s\mu_k'(s,f)\big) + O_1\left(\frac{1}{\sqrt{N}}\right)\right]\right\} \quad (5.36)$$

or

$$\begin{aligned}\frac{L}{M} &\geq F(\mathcal{Y}_m)\\ &> \exp\left\{N\left[\sum_{k=1}^{K} q_k\big(\mu_k(s,f) + (1-s)\mu_k'(s,f)\big) + O_2\left(\frac{1}{\sqrt{N}}\right)\right]\right\}\end{aligned} \quad (5.37)$$

where, for large N, the terms $O_1\left(\frac{1}{\sqrt{N}}\right)$ and $O_2\left(\frac{1}{\sqrt{N}}\right)$ scale like the inverse of the square root of N; hence, they asymptotically do not affect the exponents of these two bounds for large N. The inequalities in (5.36) and (5.37) provide a parametric lower bound on $P_{\mathrm{e,max}}$ for a given $\frac{L}{M}$ in terms of the parameter s (where $0 < s < 1$) in the same way Theorem 5.9 provides a parametric lower bound on $P_{\mathrm{e},1}$ for a given $P_{\mathrm{e},2}$. This lower bound is valid for an arbitrary fixed composition block code of composition $\mathbf{q}$ with M codewords of length N, and for an arbitrary list decoder whose lists are limited to size L. Since the lower bound is a function of an arbitrary tilting measure f which can be expanded in the form (5.30), one would in general wish to find the function f which *maximizes* the lower bound on $P_{\mathrm{e,max}}$ for a given composition. In addition, one may look for the best composition in the sense that the composition $\mathbf{q}$ *minimizes* this lower bound. This is a kind of a min-max optimization problem. The authors of [184] find the following solution for this min-max problem: For $0 < s < 1$, let $\mathbf{q}_s = (q_{1,s}, \ldots, q_{K,s})$ satisfy the condition

$$\sum_{j=1}^{J} P(j|k)^{1-s}(\alpha_{j,s})^{\frac{s}{1-s}} \geq \sum_{j=1}^{J} (\alpha_{j,s})^{\frac{1}{1-s}}\ , \quad \forall\, k = 1, \ldots, K \quad (5.38)$$

where

$$\alpha_{j,s} \triangleq \sum_{k=1}^{K} q_{k,s} P(j|k)^{1-s} \tag{5.39}$$

and let $\mathbf{f}_s \triangleq (f_s(1), \ldots, f_s(J))$ in the product form (5.30) be given by

$$f_s(j) = \frac{(\alpha_{j,s})^{\frac{1}{1-s}}}{\sum_{j'=1}^{J} (\alpha_{j',s})^{\frac{1}{1-s}}} \,. \tag{5.40}$$

Note that by multiplying both sides of (5.38) by $q_{k,s}$ and summing over $k \in \{1, \ldots, K\}$, the condition in (5.38) should be satisfied in equality for all k for which $q_{k,s} > 0$. This form of solution is reminiscent of the form of solution given in [83, Theorem 5.6.5] for the random coding error exponent. To this end, one simply needs to substitute $\rho = \frac{s}{1-s}$ in order to make Eqs. (5.38) and (5.39) look similar to those referring to the random coding error exponent in [83, Eqs. (5.6.37) and (5.6.38)]. In light of the analysis of the random coding error exponent [82], and the similarity of the two expressions, it was ensured in [184] that a solution to the equations given in (5.38) and (5.39) necessarily exists. Since the parameter s lies within the interval $[0, 1)$, then from the above substitution, the range of ρ is $[0, \infty)$; this is in contrast to the random coding error exponent where ρ is restricted to lie within the interval $[0, 1]$ (see Section 4.2.1). The methodology of the bounding technique in [184], used for the derivation of the sphere-packing bound for fixed composition codes, was to verify that unless the lower bound on $P_{\mathrm{e,max}}$ becomes trivial (i.e., the case where it is equal to zero), there exists a value of $s \in [0, 1)$ for which inequality in (5.37) is violated. Hence, the lower bound on $P_{\mathrm{e,max}}$, as given in (5.36), should be necessarily satisfied for this value of s. This approach gives the following lower bound on the maximal decoding error probability:

$$P_{\mathrm{e,max}} > \exp\left[-N\left(E_0\Big(\frac{s}{1-s}, \mathbf{q}_s\Big) - \frac{s}{1-s}\left(R - \frac{\ln 4}{N}\right) + O\Big(\frac{1}{\sqrt{N}}\Big)\right)\right] \tag{5.41}$$

where the term $O\Big(\frac{1}{\sqrt{N}}\Big)$ scales like the inverse of the square root of N and vanishes as N tends to infinity. With the substitution $\rho = \frac{s}{1-s}$, as mentioned above (so that $\rho \geq 0$), and by taking the maximization of the error exponent with respect to ρ within the interval $[0, \infty)$, the authors end up with a sphere-packing lower bound for fixed composition codes. The final form of this bound is stated in the following theorem:

Theorem 5.10. (Sphere-packing lower bound on the decoding error probability for fixed composition codes) [184, Theorem 6]: Let $P(j|k)$ be the transition probability which characterizes a DMC, and let $\mathcal{C}$ be a *fixed composition code* which is communicated over the DMC and which is comprised of M codewords of length N. Assume an arbitrary list decoding scheme where the size of the list is limited to L. Then, the *maximal error probability* is lower bounded by

$$P_{\mathrm{e,max}} \geq \exp\left[-N\left(E_{\mathrm{sp}}\left(R - \frac{\ln 4}{N} - \varepsilon\right) + \sqrt{\frac{8}{N}}\ln\left(\frac{e}{\sqrt{P_{\min}}}\right) + \frac{\ln 4}{N}\right)\right] \tag{5.42}$$

where $R \triangleq \frac{\ln\left(\frac{M}{L}\right)}{N}$, $P_{\min}$ designates the smallest non-zero transition probability of the DMC, the parameter ε is an arbitrary small positive number, and the function E_{sp} is given by

$$E_{\mathrm{sp}}(R) \triangleq \sup_{\rho \geq 0}\big(E_0(\rho) - \rho R\big) \tag{5.43}$$

$$E_0(\rho) \triangleq \max_{\mathbf{q}} E_0(\rho, \mathbf{q}) \tag{5.44}$$

$$E_0(\rho, \mathbf{q}) \triangleq -\ln\left(\sum_{j=1}^{J}\left[\sum_{k=1}^{K} q_k P(j|k)^{\frac{1}{1+\rho}}\right]^{1+\rho}\right). \tag{5.45}$$

The maximum in (5.44) is over all probability vectors $\mathbf{q} = (q_1, \ldots, q_K)$, i.e., over all $\mathbf{q}$ with K non-negative components summing to 1.

In the following, we outline the ideas which finally lead to the generalization of the sphere-packing bound for arbitrary block codes (i.e., without imposing the restriction that the codes possess a fixed composition).

Concepts used for the generalization of the SP67 bound for general block codes

The sphere-packing bound was finally generalized in [184] for block codes which do not necessarily have a fixed composition. One fact which is useful in this generalization is that the number of different ways to choose the composition of a code word is equal to the number of ways of picking K non-negative integers which sum to N. Hence, there are $\binom{N+K-1}{K-1}$ different compositions, and since the total number of codewords is equal to M, there must be some composition which contains a number of codewords M' bounded by

$$M' \geq \frac{M}{\binom{N+K-1}{K-1}} \,. \tag{5.46}$$

Since the authors in [184] are mainly focused on the analysis of the error exponent, they use the upper bound

$$\binom{N+K-1}{K-1} \leq N^K \tag{5.47}$$

for simplifying the final form of their bound; this loosening does not incur any penalty on the error exponent, as the number of different compositions only grows *polynomially* with N. Clearly, this loosening of the bound is not in place while considering a moderate value of N, as noted in [204] (see the next section which explains the improvements of the SP67 bound for moderate block lengths). By Considering the set of messages which corresponds to this set of M' codewords as a fixed composition code, and assuming that the same list decoder is used as for the original code, then for each m in this fixed composition set, $P_{\mathrm{e},m}$ stays the same as in the original code. Theorem 5.10 now applies here by replacing the rate R with $\frac{\ln\left(\frac{M'}{L}\right)}{N}$; since the error exponent E_{sp} is a decreasing function of its argument, one can loosen the lower bound by replacing M' with $\frac{M}{N^K}$, and obtain the following lower bound on the maximal error probability from Theorem 5.10:

$$P_{\mathrm{e,max}} \geq \exp\left[-N\left(E_{\mathrm{sp}}\left(\frac{\ln(\frac{M}{L})}{N} - \frac{K\ln N}{N} - \frac{\ln 4}{N}\right)\right.\right.$$
$$\left.\left. + \sqrt{\frac{8}{N}}\ln\left(\frac{e}{\sqrt{P_{\min}}}\right) + \frac{\ln 4}{N}\right)\right] \tag{5.48}$$

where the $\varepsilon > 0$ in (5.42) was chosen to absorb the inequality in (5.47).

In order to obtain a lower bound on the average error probability (rather than the maximal error probability), the authors in [184] use the concept of expurgating half of the codewords with the largest error probability, and rely on to the following simple inequality which is valid for an arbitrary block code with M codewords of length N and list decoding of size L:

$$P_{\mathrm{e}}(N,M,L) \geq \frac{1}{2}P_{\mathrm{e,max}}(N,\frac{M}{2},L).$$

This inequality relies on the fact that the block code composed from the $\frac{M}{2}$ codewords with the lowest error probability in the original code has the property that its maximal error probability does not exceed twice the average error probability of the original code (whose number of codewords is M). This finally gives the following sphere-packing lower bound:

Theorem 5.11. (1967 Sphere-packing bound for DMCs) [184, Theorem 2]: Let $\mathcal{C}$ be an arbitrary block code whose transmission takes place over a DMC. Assume that the DMC is specified by the set of transition probabilities $P(j|k)$ where $k \in \{1,\ldots,K\}$ and $j \in \{1,\ldots,J\}$ designate the channel input and output alphabets, respectively. Assume that the code $\mathcal{C}$ forms a set of M codewords of length N (i.e., each codeword is a sequence of N letters from the input alphabet), and assume that the code is decoded with an arbitrary list decoder where the size of the list is limited to L. Then, the *average decoding error probability* of the code $\mathcal{C}$ is lower bounded by

$$P_{\mathrm{e}}(N,M,L) \geq \exp\left\{-N\left[E_{\mathrm{sp}}\left(R - O_1\left(\frac{\ln N}{N}\right)\right) + O_2\left(\frac{1}{\sqrt{N}}\right)\right]\right\} \tag{5.49}$$

where $R \triangleq \frac{\ln\left(\frac{M}{L}\right)}{N}$, the error exponent $E_{\mathrm{sp}}(R)$ is introduced in (5.43), the terms

$$\begin{aligned} O_1\Big(\frac{\ln N}{N}\Big) &= \frac{\ln 8}{N} + \frac{K \ln N}{N} \\ O_2\Big(\frac{1}{\sqrt{N}}\Big) &= \sqrt{\frac{8}{N}}\, \ln\Big(\frac{e}{\sqrt{P_{\min}}}\Big) + \frac{\ln 8}{N} \end{aligned} \tag{5.50}$$

scale like $\frac{\ln N}{N}$ and the inverse of the square root of N, respectively (hence, both terms vanish as we let N tend to infinity), and $P_{\min}$ designates the smallest non-zero transition probability of the DMC.

Comparing the expression for the sphere-packing error exponent (E_{sp}) with (4.5) and (4.6), one verifies that the only difference between $E_{\mathrm{sp}}(R)$ and $E_{\mathrm{r}}(R)$ which form upper and lower bounds on the error exponents, respectively, is the range of the parameter ρ. For the upper bound, ρ can take values in the semi-infinite interval $\rho \geq 0$, where for the lower bound, ρ is restricted to the finite interval $0 \leq \rho \leq 1$. Hence, the two bounds on the error exponents coincide as long as the value of ρ which maximizes the RHS of (5.43) lies within the interval $[0,1]$. Since, the optimal parameter ρ forms a non-increasing function of the rate R, once the upper and lower bounds on the error exponent coincide at a certain rate R, they should also coincide at higher rates. Hence, the error exponents of the random coding and sphere-packing bounds coincide at the portion of the rate region between the critical rate, defined by $R_{\mathrm{c}} \triangleq E_0'(1)$, and the channel capacity. This is the rate region where the performance of capacity-approaching codes (e.g, turbo-like codes) is most appealing.

5.4 Sphere-packing bounds revisited for moderate block lengths

The analysis related to the derivation of the SP67 bound was mainly asymptotic; the purpose of the authors in [184] (see also Section 5.3) was to simplify their analysis as long as there was no penalty with respect to the error exponent of the resulting bound. It was therefore very encouraging that the error exponent of the SP67 bound [184] coincides with the error exponent of the random coding bound [82] for the whole rate region between the critical rate of a DMC and its capacity.

The introduction of turbo-like codes (e.g., turbo, LDPC and RA codes) which closely approach the channel capacity with feasible complexity (see e.g., [158] and [156]) revolutionized the field of coding theory. During the last decade, there is a large interest in the construction of structured codes with moderate block lengths which closely approach the capacity limit. Based on these exciting developments, Valembios and Fossorier [204] were motivated to review the derivation of the SP67 bound, and suggested several improvements which tighten the bound for moderate block lengths. They have also made this bound applicable for memoryless continuous-output channels (and not only for DMCs); as was exemplified in [204] (and is shown later in this section), the modified SP67 bound when particularized for the Gaussian channel competes well with the SP59 bound [185] (see Section 5.2) for practical rates and block lengths.

The derivation of the improved version of the SP67 bound in [204] deviates from the classical derivation of the SP67 bound [184] (see Section 5.3) in the following respects:

- Instead of the region $\mathcal{Y}_s$ introduced in (5.26), the derivation of the modified SP67 bound introduces an additional free parameter x, to define the set

  $$\begin{aligned}\widetilde{\mathcal{Y}}_s^x &\triangleq \left\{\mathbf{y} : |D(\mathbf{y}) - \mu'(s)| \leq x\sqrt{2\mu''(s)}\right\}\\ &= \left\{\mathbf{y} : |D(\mathbf{y}) - \mathbb{E}_{Q_s}\{D(\mathbf{y})\}| \leq x\sqrt{2\mathrm{Var}_{Q_s}\{D(\mathbf{y})\}}\right\}.\end{aligned}$$

  Then, the Chebychev inequality gives

  $$\Pr(\widetilde{\mathcal{Y}}_s^x) \geq 1 - \frac{1}{2x^2}.$$

  Clearly, this region particularizes to the region (5.26) in the specific case where $x = 1$. The lower bound above is meaningful as long as $x > \frac{\sqrt{2}}{2}$, and this generalization is used in the continuation of the derivation of the classical SP67 bound [184] (as outlined in Section 5.3).

- The simple (though loose) bound on the second derivative of the function μ in (5.35) is replaced by the exact value, expressed in terms of the transition probability of the DMC (see [204, Appendix B]). By doing so, the authors also circumvented the need for the term $\ln \frac{e}{\sqrt{P_{\min}}}$ which originally prevented the application of the classical SP67 bound to memoryless continuous-output channels due to the fact that $P_{\min}$ in Theorem 5.11 becomes infinitesimally small for memoryless continuous-output channels.
- In lieu of the upper bound (5.47) on the binomial coefficient which behaves polynomially in N, the authors in [204] use the exact value of the binomial coefficient $\binom{N+K-1}{K-1}$.
- To emphasize the similarity between the random-coding and the SP67 bounds, the parameter ρ was selected sub-optimally. It was therefore suggested in [204] to set ρ optimally in order to maximize the lower bound.

These modifications lead to the following theorem which suggests an improvement over the classical SP67 bound in terms of the pre-exponent. Note that a lower bound on the ML decoding error probability follows as a particular case of a list decoder with size $L = 1$.

Theorem 5.12. (Improvement on the 1967 sphere-packing bound for DMCs) [204, Theorem 7]: Under the assumptions and notation used in Theorem 5.11, the *average decoding error probability* is lower bounded by

$$P_{\mathrm{e}}(N,M,L) \geq \exp\Big\{-N\widetilde{E}_{\mathrm{sp}}(R,N)\Big\} \tag{5.51}$$

where

$$\widetilde{E}_{\mathrm{sp}}(R,N) \triangleq \inf_{x>\frac{\sqrt{2}}{2}} \left\{E_0(\rho_x) - \rho_x\left(R - O_1\Big(\frac{\ln N}{N}, x\Big)\right) \right. \\ \left. + O_2\Big(\frac{1}{\sqrt{N}}, x, \rho_x\Big)\right\} \tag{5.52}$$

and

$$R \triangleq \frac{\ln\left(\frac{M}{L}\right)}{N} \tag{5.53}$$

$$O_1\Big(\frac{\ln N}{N}, x\Big) \triangleq \frac{\ln 8}{N} + \frac{\ln \binom{N+K-1}{K-1}}{N} - \frac{\ln\big(2 - \frac{1}{x^2}\big)}{N} \tag{5.54}$$

$$O_2\Big(\frac{1}{\sqrt{N}}, x, \rho\Big) \triangleq x\sqrt{\frac{8}{N}\sum_{k=1}^{K} q_{k,\rho}\nu_k^{(2)}(\rho)} + \frac{\ln 8}{N} - \frac{\ln\big(2 - \frac{1}{x^2}\big)}{N} \tag{5.55}$$

$$\nu_k^{(1)}(\rho) \triangleq \frac{\sum_{j=1}^{J}\beta_{j,k,\rho}\ln\frac{\beta_{j,k,\rho}}{P(j|k)}}{\sum_{j=1}^{J}\beta_{j,k,\rho}} \tag{5.56}$$

$$\nu_k^{(2)}(\rho) \triangleq \frac{\sum_{j=1}^{J}\beta_{j,k,\rho}\ln^2\frac{\beta_{j,k,\rho}}{P(j|k)}}{\sum_{j=1}^{J}\beta_{j,k,\rho}} - \big[\nu_k^{(1)}(\rho)\big]^2 \tag{5.57}$$

$$\beta_{j,k,\rho} \triangleq P(j|k)^{\frac{1}{1+\rho}} \cdot \left(\sum_{k'} q_{k',\rho}P(j|k')^{\frac{1}{1+\rho}}\right)^{\rho} \tag{5.58}$$

where $\mathbf{q}_\rho \triangleq (q_{1,\rho}, \dots, q_{K,\rho})$ designates the input distribution which maximizes $E_0(\rho, \mathbf{q})$ in (5.44), and the parameter $\rho = \rho_x$ is determined by solving the equation

$$R - O_1\Big(\frac{\ln N}{N}, x\Big) = -\frac{1}{\rho}\sum_k q_{k,\rho}\nu_k^{(1)}(\rho) + \frac{x}{\rho}\sqrt{\frac{2}{N}\sum_{k=1}^{K} q_{k,\rho}\nu_k^{(2)}(\rho)}. \tag{5.59}$$

The proof is given in [204, pp. 3005–3006] and [204, Appendix B].

The new version of the SP67 bound in Theorem 5.12 applies to an arbitrary memoryless continuous-output channel where sums over the channel output alphabet are replaced by integrals. This bound is particularized in [204, Section 4] to the special case of a transmission over the AWGN channel with BPSK modulation. Fig. 5.2 depicts a comparison in terms of regions in the two-dimensional space (R, N) where a bound is better than the two others for three different targets (P_t) of block error probability; the three compared bounds are the improved sphere-packing bound of Valembois and Fossorier [204], the SP59 bound of Shannon and the bound which follows from the capacity limit. For example, for a rate of $\frac{3}{4}$ bits per channel use and a block error

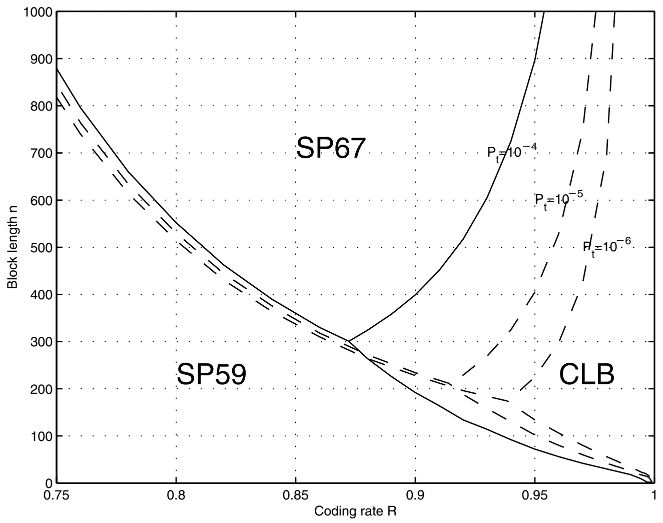

Fig. 5.2 Regions in the two-dimensional space (R, N) where one bound is better than the two others for three different targets (P_t) of block error probability. The plot corresponds to the case where the binary codewords are BPSK modulated and transmitted over the AWGN channel, and the focus is on high code rates. The SP59 and SP67 bounds refer to the sphere-packing bound of Shannon [185] and the improved version of the SP67 bound by Valembois and Fossorier [204], respectively; the CLB refers to the capacity limit bound. The figure is reproduced (with permission) from [204].

probability of 10^{-6}, the improved SP67 bound is better than the SP59 bound for block lengths above $N = 872$ bits.

5.5 Concluding comments

The introduction of structured constructions of turbo-like codes which closely approach the Shannon capacity limit with moderate block lengths stirred up new interest in studying the limits of code performance as a function of the block length. The 1959 sphere-packing bound of Shannon [185] serves for the evaluation of the performance limits of block codes over an AWGN channel where the bound is expressed in terms of the block length and code rate. This bound was used in [57] and [115] as a benchmark in order to quantify the sub-optimality of turbo codes with their practical iterative decoding algorithms; by comparing computer simulations for the performance obtained by turbo codes over a wide range of rates and block sizes, it was exemplified that the gap between their actual performance and the sphere-packing bound was within 0.7 dB. The study of Valembois and Fossorier [204] was also stimulated by the outstanding performance of turbo-like codes with moderate block lengths and feasible decoding complexity; they have improved the tightness of the classical 1967 sphere-packing bound [184] for moderate block lengths, and extended its validity to memoryless continuous-output channels (and not only for DMCs as in [184]). The improved version of the 1967 sphere-packing bound in [204] competes well with the 1959 sphere-packing bound which applies to the AWGN channel (see also Fig. 5.2), and their study stimulates further research in the adaptation of sphere-packing type lower bounds for codes of moderate block lengths which are tailored in a more suitable way for specific channel models and certain types of modulations.

In [8], the sphere-packing bound was used to assess the performance limitations of transmission techniques employing efficient block codes over fading channels. In [76], a sphere-packing bound was derived on the average decoding error probability of block codes over Rayleigh block fading multiple-input multiple-output (MIMO) channels. The results show that the performance limits, as given by sphere-packing bounds, improve significantly if a code spans a larger number of fading blocks. On the other hand, increasing the block length improves the performance limits only marginally; as we let the block length tend to infinity, the performance limits of space-time codes are determined by the outage probability (see [76]).

In [177], the authors suggest a technique to perform the entire calculation of the 1959 sphere-packing bound in the logarithmic domain. This technique eliminates the possibility of numerical overflows (see the note below Theorem 5.6 on p. 160), and it facilitates the exact calculation of the 1959 sphere-packing bound for moderate to large block lengths without the use of asymptotic approximations from [185].

6

Lower Bounds Based on de Caen's Inequality and Recent Improvements

Overview: This section introduces de Caen's lower bound, and its recent improvements by Cohen and Merhav. It provides lower bounds on the ML decoding error probability of arbitrary linear block codes which solely depend on the distance spectrum of these codes.

6.1 Introduction

The union bound asserts that the probability of a union of events does not exceed the sum of the probabilities of these events. This upper bound on the probability of a union of events forms a simple upper bound whose obvious drawback is that it may be loose due to possible intersections between the individual events. As we have seen in Section 2, for large enough codes, the union bound becomes loose at rates exceeding the channel cutoff rate, thus excluding the portion of the rate region where the outstanding performance of turbo-like codes is most appealing.

In [42], D. de Caen provides a lower bound on the probability of a finite union of events. While an elementary result (essentially, the Cauchy-Schwartz inequality), this bound was used to compute lower bounds on the decoding error probability of linear block codes via their distance spectra. As examples of such lower bounds, see [108] for the binary symmetric channel (BSC), and [182] for the binary-input AWGN channel. In [39], Cohen and Merhav improved de Caen's inequality by introducing an arbitrary non-negative weighting function which is subject to optimization. Their improved bound is presented in the following statement and, like de Caen's inequality, it follows from the Cauchy-Schwartz inequality.

Theorem 6.1. [39, Theorem 2.1] Let $\{A_i\}_{i\in\mathcal{I}}$ be a finite set of events in a probability space $(\Omega, \mathcal{F}, P)$, then the probability of the union of these events is lower bounded by

$$P\left(\bigcup_{i\in\mathcal{I}} A_i\right) \geq \sum_{i\in\mathcal{I}} \left\{ \frac{\left(\displaystyle\sum_{x\in A_i} p(x) m_i(x)\right)^2}{\displaystyle\sum_{j\in\mathcal{I}} \sum_{x\in A_i \cap A_j} p(x) m_i(x)^2} \right\} \qquad (6.1)$$

where $m_i(x) \geq 0$ is any real function on Ω such that the sums on the RHS of (6.1) converge. Further, equality in (6.1) is achieved when

$$m_i(x) = m^*(x) = \frac{1}{\deg(x)}\,, \quad \forall\, i \in \mathcal{I} \qquad (6.2)$$

where for each $x \in \Omega$

$$\deg(x) \triangleq |\{i \in \mathcal{I} \mid x \in A_i\}|. \qquad (6.3)$$

Proof. First consider the case where Ω is finite. Using a simple counting argument gives

$$P\left(\bigcup_{i\in\mathcal{I}} A_i\right) = \sum_{i\in\mathcal{I}} \sum_{x\in A_i} \frac{p(x)}{\deg(x)}. \qquad (6.4)$$

Let $m_i(x) \geq 0$ be an arbitrary non-negative function on Ω. From the Cauchy-Schwarz inequality, it follows that

$$\left(\sum_{x \in A_i} \frac{p(x)}{\deg(x)}\right)\left(\sum_{x \in A_i} p(x) m_i^2(x) \deg(x)\right)$$
$$\geq \left(\sum_{x \in A_i} \sqrt{\frac{p(x)}{\deg(x)}} \sqrt{p(x) m_i^2(x) \deg(x)}\right)^2$$
$$= \left(\sum_{x \in A_i} p(x) m_i(x)\right)^2 \tag{6.5}$$

provided that the sums in (6.5) converge. Therefore, (6.4) and (6.5) give

$$P\left(\bigcup_{i \in \mathcal{I}} A_i\right) \geq \sum_{i \in \mathcal{I}} \frac{\left(\sum_{x \in A_i} p(x) m_i(x)\right)^2}{\sum_{x \in A_i} p(x) m_i^2(x) \deg(x)}$$
$$= \sum_{i \in \mathcal{I}} \frac{\left(\sum_{x \in A_i} p(x) m_i(x)\right)^2}{\sum_{j \in \mathcal{I}} \sum_{x \in A_i \cap A_j} p(x) m_i^2(x)}. \tag{6.6}$$

Note that $m_i(x)$ may be different for each i in the sum over all $i \in \mathcal{I}$. However, in order to achieve equality in (6.5), we need (6.2) to be satisfied.

For a general probability space, as noted in [42] and [113], since there are only finitely many A_i's, the number of Boolean atoms defined by the A_i's unions and intersections is also finite. Thus, the general space can be reduced to a finite probability space. In this case, the sums in (6.1) are replaced by integrals. □

In [39], Cohen and Merhav refer to the choice of $m_i(x) \equiv 1$ as the *trivial* choice of $m_i(x)$. In this particular case, we get the inequality

$$P\left(\bigcup_{i \in \mathcal{I}} A_i\right) \geq \sum_{i \in \mathcal{I}} \left\{ \frac{P(A_i)^2}{\sum_{j \in \mathcal{I}} P(A_i \cap A_j)} \right\}$$

which is de Caen's bound [42]. Thus, de Caen's bound is a special case of the bound suggested by Cohen and Merhav in Theorem 6.1. In this context, it is noted that a recent improvement of de Caen's bound was provided by Kuai, Alajaji and Takahara [113]. However, Dembo has demonstrated in [44] that both de Caen's bound and Kuai, Alajaji and Takahara's bound are derived by solving the same minimization problem. While the latter bound is obtained by applying a stronger method than de Caen's bound, it improves this bound by a negligible factor which is at most $\frac{9}{8}$ (see [44]).

The essence of the bound in Theorem 6.1 is the ability to choose an appropriate function $m_i(x)$. To define a proper strategy for choosing $m_i(x)$, it was observed by Cohen and Merhav that a multiplication of this weighting function by a constant factors out in (6.1). Hence, $m_i(x)$ should only define an *essence of behavior*, and not necessarily exact values. When looking for such a behavior, recall that the optimal function is given by $m_i(x) = \frac{1}{\deg(x)}$ (see (6.2)). However, in the context of the derivation of lower bounds on the ML decoding error probability for linear block codes, the calculation of this function is prohibitively complex, as it requires the full characterization of the Voronoi regions of every individual codeword. For this reason, several alternatives are suggested in [39] for choosing the functions $m_i(x)$; these choices yield inherently tighter bounds than some previously reported lower bounds which rely on de Caen's inequality (e.g., Seguin's bound [182]). By following this approach, Cohen and Merhav relied on (6.1) for the derivation of improved lower bounds on the decoding error probability of linear codes under optimal ML decoding. They exemplified their bounds for BPSK modulated signals which are equally likely to be transmitted among M signals, and where the examined communication channels are a BSC or a binary-input AWGN channel. In this context, the element x in the RHS of (6.1) is replaced by the received vector $\mathbf{y}$ at the output of the communication channel, and A_i (where $i = 1, 2, \ldots, M-1$) consists of all the vectors which are closer in the Euclidean sense to the signal $\mathbf{s}_i$ rather than the transmitted signal $\mathbf{s}_0$. Similarly to Seguin's bound [182], the bounds in [39] get (after some loosening in their tightness) final forms which solely depend on the distance spectrum of the code. Recently, two lower bounds on the ML decoding error probability of binary linear block codes were derived by Behnamfar et al. [16] for BPSK-modulated AWGN channels. These bounds are easier for numerical calculation, but are looser than Cohen-Merhav bounds for low to moderate SNRs.

We note that lower bounds which are based on de Caen's inequality (see [16], [39], [108] and [182]) are applicable for *specific* codes but not for ensembles; this restriction is due to the fact that Jensen's inequality does not allow to replace the distance spectrum of a linear code in these bounds by the average distance spectrum of ensembles.

6.2 Lower bounds based on de Caen's inequality and variations

6.2.1 Lower bounds on the ML decoding error probability for the AWGN channel

We start our discussion by presenting the lower bound of Cohen and Merhav, as introduced in [39, Proposition 3.1] for the Gaussian channel, and then refer to its two particular cases which are named in [39, Section 3] as the 'norm' and 'dot-product' lower bounds; these two lower bounds on the ML decoding error probability refer to the situation where the signals are transmitted over the AWGN channel with equal probability. These bounds rely on particular choices of the 'weighting functions' m_i which appear in Theorem 6.1. The following lower bound, introduced in Theorem 6.2, relies on the improvement of de Caen's lower bound (see Theorem 6.1). To this end, the authors make use of the following parametric weighting function, given the transmitted signal $\mathbf{s}_0$ and the received vector $\mathbf{y}$:

$$m(\mathbf{y}|\mathbf{s}_0) = \exp\left\{-(a||\mathbf{y}||^2 + b\,\langle \mathbf{y}, \mathbf{s}_0\rangle + c||\mathbf{s}_0||^2)\right\} \tag{6.7}$$

where a, b and c are free parameters, and $\langle\cdot,\cdot\rangle$ designates the standard inner product. It is easily demonstrated that Seguin's bound [182] forms a particular case of the lower bounds of Cohen and Merhav, and therefore Seguin's bound is in general a looser bound (as exemplified in [39] and in some numerical results in this section). In order to derive lower bounds which solely depend on the distance spectrum of a binary linear block code, Seguin proved a monotonicity property with respect to the correlation between two arbitrary codewords of the code. This property enables to loosen the lower bounds by replacing the correlation between two arbitrary codewords with a simple upper bound which depends on the Hamming weights of the two codewords and the minimum distance of the code. The same upper bound on the correlation between two arbitrary codewords was also used in [39]. We finally address Swatzek's bound [192] which was exemplified in [182] to be looser than the bound of Seguin. The discussion concludes by addressing a recently introduced bound of Benhamfar, Alajaji and Linder [16]. Some numerical results finally exemplify the use of these distance-spectrum based bounds for the performance evaluation of some (short) block codes; these bounds are also compared with the 1959 sphere-packing (SP59) bound of Shannon (see Section 5.2) and the TSB upper bound of Poltyrev (see Section 3.2.1).

Theorem 6.2. [39, Proposition 3.1] Let $\mathbf{s}_0, \ldots, \mathbf{s}_{M-1}$ be a set of M signals of dimension K whose transmission takes place over an AWGN channel with two-sided noise spectral density of $\frac{N_0}{2}$. Then, the conditional error probability under ML decoding given that the signal $\mathbf{s}_0$ is transmitted is lower bounded by

$$P(\text{error}|\mathbf{s}_0) \geq \exp\bigl((\beta' - 2\beta)||\mathbf{s}_0||^2\bigr) \left(\frac{N_0'}{\sqrt{N_0 N_0''}}\right)^K \sum_{i=1}^{M-1} \left\{ \frac{Q^2\bigl(\kappa(\alpha, \mathbf{s}_i, N_0')\bigr)}{\displaystyle\sum_{j=1}^{M-1} \Psi\bigl(\rho_{i,j}, \kappa(\alpha', \mathbf{s}_i, N_0''), \kappa(\alpha', \mathbf{s}_j, N_0'')\bigr)} \right\} \tag{6.8}$$

where

$$Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\bigl(-\frac{y^2}{2}\bigr)\, dy \tag{6.9}$$

is the complementary Gaussian cumulative distribution function,

$$\Psi(\rho, x', y') \triangleq \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{x'}^\infty \int_{y'}^\infty \exp\left\{-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right\} dx\, dy \tag{6.10}$$

is the normalized (unit variance) two-dimensional Gaussian integral with $\Psi(1, x, x) \triangleq Q(x)$,

$$\rho_{i,j} \triangleq \frac{\langle \mathbf{s}_i - \mathbf{s}_0, \mathbf{s}_j - \mathbf{s}_0 \rangle}{||\mathbf{s}_i - \mathbf{s}_0||\; ||\mathbf{s}_j - \mathbf{s}_0||} \tag{6.11}$$

designates a correlation, the function κ is given by

$$\kappa(\alpha,\mathbf{s}_i,N_0) \triangleq \frac{||\alpha\mathbf{s}_0-\mathbf{s}_i||^2-(\alpha-1)^2||\mathbf{s}_0||^2}{\sqrt{2N_0}\,||\mathbf{s}_i-\mathbf{s}_0||}.$$

The constants $N_0', N_0'', \alpha, \alpha', \beta$ and β' are given by

$$N_0' \triangleq \tfrac{N_0}{1+aN_0}\,, \quad N_0'' \triangleq \tfrac{N_0}{1+2aN_0}$$

$$\alpha \triangleq \tfrac{\frac{1}{N_0}-\frac{b}{2}}{a+\frac{1}{N_0}}\,, \quad \alpha' \triangleq \tfrac{\frac{1}{N_0}-b}{2a+\frac{1}{N_0}}$$

$$\beta \triangleq \tfrac{\left(\frac{1}{N_0}+a\right)\left(\frac{1}{N_0}+c\right)-\left(\frac{1}{N_0}-\frac{b}{2}\right)^2}{\frac{1}{N_0}+a}\,,$$

$$\beta' \triangleq \tfrac{\left(\frac{1}{N_0}+2a\right)\left(\frac{1}{N_0}+2c\right)-\left(\frac{1}{N_0}-b\right)^2}{\frac{1}{N_0}+2a} \tag{6.12}$$

where $a > -\frac{1}{2N_0}$, and the constants b and c are arbitrary real numbers.

For ease of computation of the lower bound in Theorem 6.2, it is useful to rely on the following alternative representations of the functions Q and Ψ from Eqs. (6.9) and (6.10)

$$Q(x) \triangleq \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \exp\left(-\frac{x^2}{2\sin^2\theta}\right)d\theta\,, \quad x > 0 \tag{6.13}$$

and

$$\begin{aligned}
&\Psi(\rho,x,y)\\
&= \frac{1}{2\pi}\int_0^{\frac{\pi}{2}-\tan^{-1}\left(\frac{y}{x}\right)} \frac{\sqrt{1-\rho^2}}{1-\rho\sin 2\theta}\exp\left(-\frac{x^2}{2}\frac{1-\rho\sin 2\theta}{(1-\rho^2)\sin^2\theta}\right)d\theta\\
&\quad + \frac{1}{2\pi}\int_0^{\tan^{-1}\left(\frac{y}{x}\right)} \frac{\sqrt{1-\rho^2}}{1-\rho\sin 2\theta}\exp\left(-\frac{y^2}{2}\frac{1-\rho\sin 2\theta}{(1-\rho^2)\sin^2\theta}\right)d\theta
\end{aligned} \tag{6.14}$$

where Eq. (6.13) was first introduced by Craig [40], and Eq. (6.14) was derived by Simon and Divsalar (see [188, Section 9]).

Let $\mathcal{C}$ be a binary linear block code whose codewords are BPSK modulated, and transmitted over an AWGN channel. Assume that the only information we have about the linear code $\mathcal{C}$ is its distance spectrum, and we are looking for a lower bound on the error probability of this code under ML decoding. Based on the linearity of the code, one can assume that the all-zero codeword is transmitted, and apply the lower bound in Theorem 6.2 to obtain a lower bound on the average decoding error probability. However, the only obstacle in applying the bound in (6.8) to linear block codes is the fact that the correlation two signals (i.e., $\rho_{i,j}$ in (6.11)) does not depend solely on the Hamming weight of the individual codewords, but also on the number of agreements (and disagreements) between these two codewords. A pleasing property of the two-dimensional Gaussian integral $\Psi(\rho,\cdot,\cdot)$ in (6.10) is that this function is a monotonically increasing function of the correlation ρ (this property was formally proved by Seguin in [182, Eqs. (23)–(25)]); hence, as suggested by Seguin, one can replace the correlation $\rho_{i,j}$ in the RHS of (6.8) by an *upper bound on this correlation* which only depends on the Hamming weights of the individual codewords, and in this way obtain a looser lower bound on the decoding error probability which solely depends on the distance spectrum of the code. To this end, Seguin relied on the following upper bound on the correlation:

$$\rho_{i,j} \leq \rho(i,j) \triangleq \min\left\{\sqrt{\frac{w_{\mathrm{H}}(\mathbf{c}_i)}{w_{\mathrm{H}}(\mathbf{c}_j)}}, \sqrt{\frac{w_{\mathrm{H}}(\mathbf{c}_j)}{w_{\mathrm{H}}(\mathbf{c}_i)}}, \frac{w_{\mathrm{H}}(\mathbf{c}_i) + w_{\mathrm{H}}(\mathbf{c}_j) - d_{\min}}{2\sqrt{w_{\mathrm{H}}(\mathbf{c}_i) w_{\mathrm{H}}(\mathbf{c}_j)}}\right\} \tag{6.15}$$

where $w_{\mathrm{H}}(\mathbf{c})$ designates the Hamming weight of a codeword $\mathbf{c}$, and $d_{\min}$ designates the minimal distance of the code $\mathcal{C}$. In [39], Cohen and Merhav relied on the same upper bound on the correlation in order to obtain a looser lower bound on the ML decoding error probability which solely depends on the distance spectrum of a binary linear block code; to this end, the correlation $\rho_{i,j}$ in the RHS of (6.8) is replaced by the upper bound on $\rho_{i,j}$, as given in (6.15), and obtain an overall lower bound which solely depends on the distance spectrum of $\mathcal{C}$.

From the statement in Theorem 6.1 and the context of Theorem 6.2 which addresses a hypothesis testing problem for the AWGN channel, the corresponding optimized weighting function m_i which maximizes the lower bound in the RHS of (6.6) is independent of the index i, and it scales like the inverse of the number of signals which are closer to the received vector $\mathbf{y}$ than the transmitted signal $\mathbf{s}_0$, i.e.,

$$m_{\mathrm{opt}}(\mathbf{y}|\mathbf{s}_0) = \frac{1}{\deg(\mathbf{y}|\mathbf{s}_0)} \tag{6.16}$$

where

$$\deg(\mathbf{y}|\mathbf{s}_0) \triangleq \big|\{i : ||\mathbf{y} - \mathbf{s}_i|| < ||\mathbf{y} - \mathbf{s}_0||\}\big| \tag{6.17}$$

Based on the three-parameter weighting function introduced in (6.7), [39] considers the following three particular cases of the bound in Theorem 6.2:

- *Norm Lower bound*: This lower bound follows by looking at the weighting functions whose behavior is like the exponent of the squared Euclidean distance between the received vector $\mathbf{y}$ and the signal $\mathbf{s}_0$. The motivation for considering this case is because, according to the optimized weighting function given in (6.16), it is inversely proportional to the number of signals which fall in the interior of a K-dimensional sphere of radius $||\mathbf{y} - \mathbf{s}_0||$ whose center is at the received vector $\mathbf{y}$, so the optimized function is monotonically decreasing in the distance $||\mathbf{y} - \mathbf{s}_0||$. Hence, Cohen and Merhav considered the one-parameter weighting function

  $$m(\mathbf{y}|\mathbf{s}_0) \triangleq \exp(-a||\mathbf{y} - \mathbf{s}_0||^2)\ , \quad a \geq 0 \tag{6.18}$$

  which follows as a particular case of the three-parameter weighting function in (6.7) when $b = -2a$ and $a = c$. The optimized bound which corresponds to this particular case is called the 'norm bound', and it follows as a particular case of the lower bound in Theorem 6.2 in the above setting of the parameters a, b and c (so that the optimization problem is reduced to a single-parameter optimization with respect to $a \geq 0$).
- *Dot-Product Lower Bound*: For equal-energy signals, the degree function in (6.17) can be rewritten in the form

  $$\begin{aligned}\deg(\mathbf{y}|\mathbf{s}_0) &= \big|\{i :\ \langle \mathbf{y}, \mathbf{s}_i\rangle\ >\ \langle \mathbf{y}, \mathbf{s}_0\rangle\}\big| \\ &= \big|\{i : \theta_{\mathbf{y},i} < \theta_{\mathbf{y},0}\}\big|\end{aligned} \tag{6.19}$$

where $\langle f, g\rangle$ designates the standard inner product of the two functions f and g, and

$$\theta_{\mathbf{y},i} \triangleq \cos^{-1}\left(\frac{\langle \mathbf{s}_i, \mathbf{y}\rangle}{||\mathbf{s}_i||||\mathbf{y}||}\right), \quad 0 \le \theta_{\mathbf{y},i} < \pi.$$

Had the equal-energy signals been uniformly distributed on the surface of a sphere centered at the origin (this assumption holds for the optimal code, and gives the 1959 sphere-packing bound of Shannon [185]), Eqs. (6.16) and (6.19) would imply that the optimized weighting function $m(\mathbf{y}|\mathbf{s}_0)$ is monotonically decreasing with respect to the absolute value of the angle between $\mathbf{y}$ and $\mathbf{s}_0$. This intuition suggested Cohen and Merhav to consider the particular weighting function

$$m(\mathbf{y}|\mathbf{s}_0) \triangleq \exp\bigl(b\,\langle \mathbf{s}_0, \mathbf{y}\rangle\bigr)\ , \quad b \in \mathbb{R} \qquad (6.20)$$

which forms a particular case of the three-parameter function in (6.7) when $a = c = 0$ and b is replaced by $-b$. Hence, the dot-product lower bound is also subject to a single-parameter optimization (with respect to the parameter b).

- *Seguin's Lower Bound*: Seguin's lower bound on the ML decoding error probability was introduced in [182], and it follows as a particular case of the Cohen-Merhav bound in Theorem 6.2 by setting $a = b = c = 0$. The reason for this is that Seguin's lower bound follows from de Caen's bound, where the latter forms a particular case of the improvement on de Caen's lower bound (see Theorem 6.1) in the setting where all the weighting functions m_i are equal to unity. Hence, by substituting $a = b = c = 0$ in the RHS of (6.7), the weighting function $m(\mathbf{y}|\mathbf{s}_0)$ is identically equal to 1. Note that Seguin's bound is also a particular case of both the norm bound and the dot-product bound.

Based on the particular choice of the weighting function in (6.7) and the upper bound on the correlation $\rho_{i,j}$ as given in (6.15), Theorem 6.2 particularizes to [39, Eqs. (24) and (25)]. However, we rewrite these equations in a different form which has two advantages: the first is that the lower bound is explicitly expressed in terms of the signal to noise ratio (as opposed to the constants in (6.12) which depend on the noise spectral density N_0 and the signal energy per symbol E_{s} separately). The second advantage is that it shows explicitly that the arbitrary parameter c in (6.7) does not affect the bound (this can be also seen directly from (6.7), since the parameter c just scales this function); hence, the calculation of the lower bound is reduced to an optimization over two parameters. The following corollary rephrases [39, Eqs. (24) and (25)] in a more convenient form, as explained above.

Theorem 6.3. (Lower bound on the decoding error probability of binary linear block codes over an AWGN channel) Let $\mathcal{C}$ be a binary linear block code of length N and dimension K, $R \triangleq \frac{K}{N}$ be the code rate of $\mathcal{C}$ in bits per channel use, $d_{\min}$ be the minimal distance of $\mathcal{C}$, and let $\{S_i\}_{i=0}^{N}$ designate the distance spectrum of $\mathcal{C}$. Assume that the code is BPSK modulated, and transmitted over an AWGN channel. Then, the block error probability of the code $\mathcal{C}$ under ML decoding is lower bounded by

$$P_{\mathrm{e}}(\mathcal{C}) \geq \exp\left\{-\frac{NRE_{\mathrm{b}}}{N_0}\left[1+\frac{(1-b)^2}{1+2a}-\frac{(2-b)^2}{2(1+a)}\right]\right\}\left(\frac{\sqrt{1+2a}}{1+a}\right)^{K}$$
$$\cdot \sum_{i\neq 0: S_i>0} \frac{S_i\, Q^2(\gamma_i)}{Q(\delta_i)+(S_i-1)\Psi\big(\rho(i,i),\delta_i,\delta_i\big)+\sum\limits_{j\neq\{0,i\}} S_j \Psi\big(\rho(i,j),\delta_i,\delta_j\big)} \tag{6.21}$$

where

$$\gamma_i \triangleq \sqrt{\tfrac{(2-b)^2}{2(1+a)}\,\tfrac{iRE_{\mathrm{b}}}{N_0}} \tag{6.22}$$

$$\delta_i \triangleq \sqrt{\tfrac{2(1-b)^2}{1+2a}\,\tfrac{iRE_{\mathrm{b}}}{N_0}} \tag{6.23}$$

$$\rho(i,j) \triangleq \min\left\{\sqrt{\tfrac{i}{j}}, \sqrt{\tfrac{j}{i}}, \tfrac{i+j-d_{\min}}{2\sqrt{ij}}\right\} \tag{6.24}$$

and $a > -\frac{1}{2}$, $b \in \mathbb{R}$ are two arbitrary constants which are subject to optimization. The functions Q and Ψ in the RHS of (6.21) are given in (6.9) and (6.10), respectively (these functions are also expressed in (6.13) and (6.14) as integrals over finite intervals.)

As explained above and since the parameter c in the weighting function (6.7), the following three bounds follow as particular cases of the bound given in Theorem 6.3:

- By setting $b = -2a$, $a \geq 0$ and optimizing over a in (6.21), one obtains the 'norm bound' [39].
- By setting $a = 0$ and optimizing over b in (6.21), one obtains the 'dot-product bound' [39]
- By setting $a = b = 0$ in (6.21) (which renders the weighting function equal to 1), one obtains Seguin's bound [182].

A comparison of the 'norm bound', the 'dot-product bound' [39] and Seguin's bound [182] which follow as particular cases of Theorem 6.3, and the 1959 sphere-packing lower bound [185] (see Section 5.2) is shown in Fig. 6.1 for two short linear block codes (the upper plot refers to the (63, 24) BCH codes, and the lower plot refers to the (24, 12) Golay code). As expected from the discussion above, Seguin's bound is looser than the 'norm bound' and the 'dot-product' bound. In order to examine the tightness of these lower bounds, they are compared with the tangential-sphere bound of Poltyrev [152] which is one of the tightest reported upper bounds for the AWGN channel (see Section 3.2.1).

The sensitivity of the 'norm bound' to the parameter a is exemplified in Fig. 6.2 for the $(23, 12, 7)$ Golay code. This figure shows a three dimensional mesh of the lower bound on the decoding error probability of the Golay code as a function of the energy per bit to spectral noise density $\left(\frac{E_\text{b}}{N_0}\right)$ and the parameter $a > -\frac{1}{2}$ of the lower bound in (6.21) which is subject to optimization.

In [16], new lower bounds on the ML decoding error probability of binary linear block codes over the Gaussian channel were recently introduced. They suggest a reduction in the computational complexity as compared to the Cohen-Merhav bound, and are therefore easy for calculation; however, they are less tight than the Cohen and Merhav bounds for low values of signal to noise ratio.

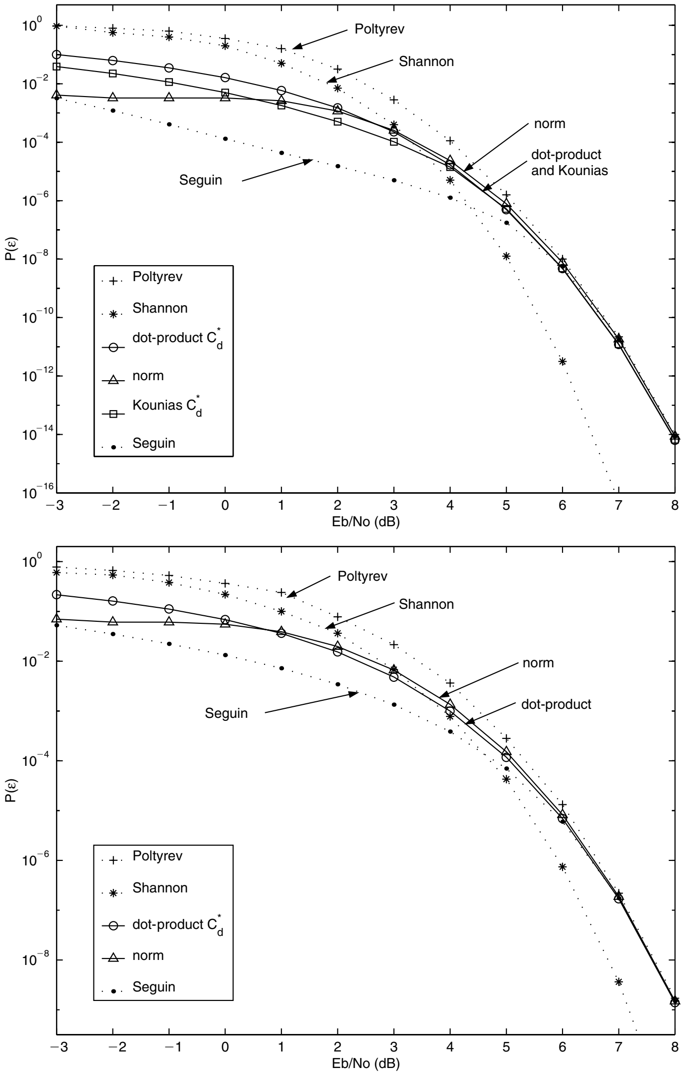

Fig. 6.1 Bounds on the decoding error probability of two binary linear block codes under soft-decision ML decoding. The upper plot refers to the (63, 24) BCH code, and the lower plot refers to the (24, 12) Golay code where the transmission of both codes takes place over the binary-input AWGN channel. Poltyrev's upper bound [152] refers to the tangential-sphere bound (TSB) in Section 3.2.1; the other bounds which are lower bounds on the ML decoding error probability include Seguin's bound [182], the 'dot-product bound' and the 'norm bound' introduced by Cohen and Merhav [39], and the 1959 sphere-packing bound (SP59) of Shannon [185] which is introduced in Section 5.2 (where the latter bound solely depends on the block length and the rate of the code). This figure was reproduced (with permission) from [39].

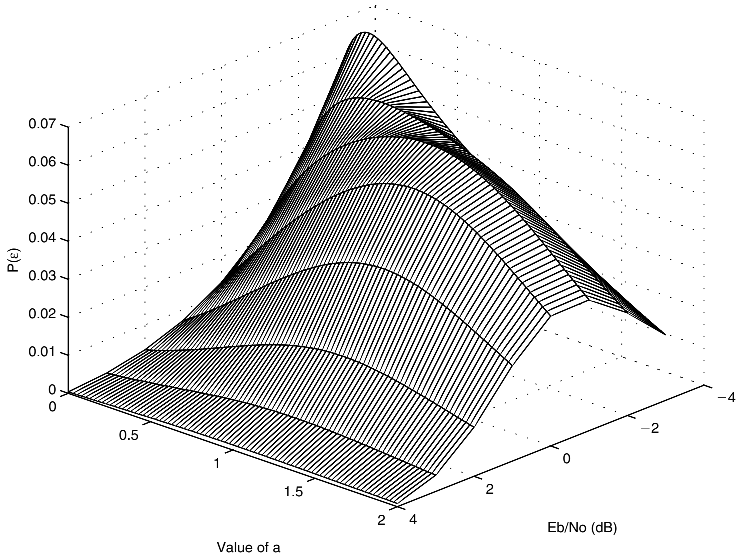

Fig. 6.2 A three-dimensional mesh of the 'norm lower bound' on the decoding error probability of the (23, 12, 7) Golay code under ML decoding. The graph of this lower bound is shown as a function of the energy per bit to spectral noise density $\left(\frac{E_{\mathrm{b}}}{N_0}\right)$ and the parameter $a > -\frac{1}{2}$ of the lower bound in (6.21) which is subject to optimization. This figure was reproduced (with permission) from [39].

6.2.2 Lower bounds for the binary symmetric channel

We discuss in this sub-section lower bounds on the ML decoding error probability of binary linear block codes over the BSC, and refer to the Keren and Litsyn bound [108] and to a bound of Cohen and Merhav which was derived for the BSC [39].

In this case, the transmitted codeword is one of $M = 2^K$ equiprobable codewords $\mathbf{c}_0, \ldots, \mathbf{c}_{M-1}$ of length N whose transmission takes place over a binary symmetric channel (BSC) with crossover probability p (without any loss of generality, one can assume that $p < \frac{1}{2}$). Let $\mathbf{c}_0$ be the transmitted codeword, then the ML decoder chooses the codeword whose Hamming distance from the received vector is minimal. Let $\mathbf{x} = \mathbf{c}_0 + \mathbf{e}$ be the received vector where $\mathbf{e} \in \mathrm{GF}(2)^N$ is the error vector. Assume that the code $\mathcal{C}$ is a binary linear block code, then the symmetry of the channel yields that the average decoding error probability is equal to the conditional error probability given that the all-zero codeword is transmitted. The error event $\mathcal{E}$ can therefore be expressed as a union of events $\mathcal{E} = \bigcup_{i=1}^{M-1} \mathcal{E}_{0,i}$ (which typically have overlaps, hence the weakness of the union bound for bad channel conditions) where

$$\mathcal{E}_{0,i} \triangleq \left\{ \mathbf{x} \in \mathrm{GF}(2)^N : \; w_{\mathrm{H}}(\mathbf{x} + \mathbf{c}_i) < w_{\mathrm{H}}(\mathbf{x}) \right\}. \tag{6.25}$$

In [39, Section 4], Cohen and Merhav derive a lower bound on the decoding error probability for the BSC which solely depends on the distance spectrum of the binary linear block code and the crossover probability p of the BSC. Similarly to the derivation of lower bounds for the AWGN channel (see Section 6.2.1), they rely on the improvement on de Caen's lower bound as stated in Theorem 6.1. In order to simplify the final form of their bound and make it expressible in terms of the distance spectrum, they chose the setting where the non-negative weighting function $m_i(\mathbf{x})$ which appears in the RHS of (6.1) depends on the received vector $\mathbf{x}$ only through its Hamming weight (so, for all vectors $\mathbf{x} \in \mathrm{GF}(2)^N$ of the same composition, the function $m_i(\mathbf{x})$ is assumed to get the same value). For the BSC model, the transition probability is given by

$$\Pr(\mathbf{x}|\mathbf{c}_0 = \mathbf{0}) = p^{w_{\mathrm{H}}(\mathbf{x})}(1-p)^{N - w_{\mathrm{H}}(\mathbf{x})}.$$

From (6.3) (where in the context of this testing hypothesis problem, the event $\mathcal{A}_i$ appearing in Theorem 6.1 is replaced with the pairwise event event $\mathcal{E}_{0,i}$ as given in (6.25)), the degree function can be expressed as follows:

$$\begin{aligned} \deg(\mathbf{x}|\mathbf{c}_0) &= \left| \{ \mathbf{c}_i \in \mathcal{C}, i \neq 0 : w_{\mathrm{H}}(\mathbf{x} + \mathbf{c}_i) < w_{\mathrm{H}}(\mathbf{x}) \} \right| \\ &= |\mathcal{C}| \; \Pr\{ w_{\mathrm{H}}(\mathbf{x} + \mathbf{c}) < w_{\mathrm{H}}(\mathbf{x}) \} \end{aligned} \tag{6.26}$$

where a uniform distribution over the codewords of $\mathcal{C}$ is assumed. From the statement in Theorem 6.1, the optimal weighting function $m_i(\mathbf{x})$ is independent of the index i, and it is inversely proportional to the value of $\deg(\mathbf{x}|\mathbf{c}_0)$. In the continuation of the discussion in [39], two alternatives for this optimized weighting function are suggested; both are amenable to analysis and yield in their final form lower bounds which solely depend on the distance spectrum of the code and the crossover probability of the channel. Since the weighting function m_i is assumed in [39, Section 4] to depend only on the Hamming weight of the received vector $\mathbf{x}$, this function is given in the form

$$m_i(\mathbf{x}) = \mu_i\big(w_{\mathrm{H}}(\mathbf{x})\big), \quad \mu_i : \mathbb{N}^+ \to \mathbb{R}.$$

The reader is referred to [39, Section 4] for further details regarding the derivation of the lower bound for the BSC (where the concept of the derivation is similar to the derivation of the lower bounds for the Gaussian channel in Section 6.2.1; it relies on the lower bound on a union of events, as given in Theorem 6.1). To this end, Cohen and Merhav suggest two functions μ_i which are based on certain approximations of the function $\deg(\mathbf{x}|\mathbf{c}_0)$; the calculation of one of these approximations depends on the knowledge of the distance spectrum of the whole code, and it is obtained from the Chernoff bounding technique; the other approximation depends on more elementary properties of the code (e.g., its size, length and minimum distance). In the final from of these bounds, both of them are subject to a one-parameter optimization in order to obtain the tightest lower bound within the considered family.

The lower bound of Keren and Litsyn [108] also applies to binary linear block codes whose transmission takes place over the BSC; however, since the derivation of their bound follows from de Caen's bound (i.e., the weighting functions in the RHS of (6.1) are all equal to unity), the bound in [108] is looser than the bound of Cohen and Merhav. Both bounds are calculable in terms of the distance spectrum of the code, and are exemplified for some class of codes in [39].

6.3 Summary and conclusions

This section introduces de Caen's lower bound on the probability of a union of events [42], and its recent improvement by Cohen and Merhav [39]. These bounds provide lower bounds on the ML decoding error probability of binary linear block codes, which solely depend on their distance spectrum. The improved bounds in [39] require an optimization over an arbitrary non-negative weighting function; the optimal choice of this function is known, but it unfortunately leads to a useless identity. Several sub-optimal choices of this function with some free parameters lead to lower bounds on the ML decoding error probability which are subject to parameter-optimizations (in order to get the tightest lower bounds within their forms). For binary linear block codes whose transmission takes place over an AWGN channel, the bound of Cohen and Merhav is reformulated in this section so that it is explicitly expressed in terms of the energy per bit to spectral noise density of the channel and the distance spectrum of the code (see Theorem 6.3). The latter form of a two-parameter bound is uniformly tighter than the 'norm bound' and the 'dot-product bound' which are studied in [39]. Finally, lower bounds for the BSC [39, 108] are shortly addressed in this section.

7

Concluding Remarks

Maximum-likelihood (ML) decoding provides an upper limit on the attainable performance of coded communication systems, and hence is of interest, even when the actual optimal decoding procedure can not be practically implemented. The implications of understanding the performance and operation of ML decoding range beyond this obvious aspect. Namely, as of late intimate relations between the ML and sub-optimal iterative message-passing algorithms have been revealed [128, 129, 130, 157], and a geometric view, advocated here in terms of different bounding techniques might be beneficial.

As an example for the possible relations between ML and iterative decoding algorithms, we note that some recent publications are focused on the tradeoff between performance (even under optimal ML decoding) and the decoding complexity per iteration for low-density parity-check (LDPC) codes (or general binary linear block codes which are iteratively decoded based on their bipartite graph) [33, 149, 176, 179, 215, 214]. This tradeoff between performance and complexity is expressed in terms of the gap (in rate) to capacity. The study of the tradeoff between achievable rates under ML decoding and the decoding complexity per iteration is done via information-theoretic bounds which also enable to get an indication as to the sub-optimality of iterative message-passing decoding algorithms (as compared to optimal ML decoding). To this end, bounds on the thresholds under ML decoding are compared with exact thresholds under iterative message-passing decoding (whose calculation is based on the density evolution technique). A generalization of the bounds for parallel channels [180, 178] enables to apply the new bounds to ensembles of punctured LDPC codes, where both intentional and random puncturing are addressed. This suggests an interesting direction for further research on the relation between performance bounds under ML decoding and the decoding complexity per iteration (under iterative message-passing decoding); this relation is based on information-theoretic bounds which are valid for every code (and not only for ensembles, via concentration arguments). Bounds of the type considered here such as variants of the Shulman and Feder bound [187] were found useful in addressing these questions (see, e.g., [94, Theorem 2], [95, Theorem 1], and [176, Theorem 2.2]).

An additional aspect which is of great interest in the context of iterative decoding and is related to performance bounds under ML decoding concerns the fundamental limitations of codes with cycles. It is well known that codes whose Tanner graphs are cycle-free have poor performance, even under optimal ML decoding [69]. In order to enhance this result, information-theoretic bounds which are valid under optimal ML decoding (or hence, for any other sub-optimal decoding algorithm) enable to relate the performance of any binary linear code defined on graph to the minimal number of fundamental cycles which should exist in an arbitrary bipartite graph characterizing this code [176, 175, 180]. This approach also allows one to derive lower bounds on the bit error probability under ML decoding as a function of the density of an arbitrary parity-check matrix which represents the code. These information-theoretic arguments explain why good codes should have cycles in their bipartite graphs, and show that this property is also valid under optimal ML decoding (so it is not a consequence of the sub-optimality of iterative message-passing algorithms). We believe that this direction deserves further study, and performance bounds under ML decoding appear to be informative also in this respect.

Yet another interesting aspect which deserves attention is the assessment of the performance of modern families of codes via statistical physics methods (see [77], [135]–[136], [157], [189] and references therein). It is most intriguing to see whether there are any conceptual connections between this promising, though not yet fully rigorous methodology, and bounding techniques of the generalized Gallager family. Evidently, any such insight could benefit significantly both domains.

The proposed approach for the derivation of Gallager bounds and their variations, as discussed in Section 4, can be generalized to geometrically uniform non-binary codes, finite-state channels, bit interleaved coded modulation systems, coding for independent parallel channels [122, 166, 165], and it can be also used for the derivation of upper bounds on the conditional decoding error probability (as to account for a possible partitioning of the original code to subcodes).

While the power and conceptual importance of Gallager's bounding methodology (including the 1961 Gallager and 1965 Gallager bounds) has been widely exploited, we still believe that this has not exhausted in full the power of Gallager-oriented approaches. Specifically, fundamental questions such as what is the deviation in some appropriate 'distance' measure (see for example (4.39), or the divergence measure used in [288, Appendix A]) of a distance spectrum of a linear code as compared to the binomial reference, that still permits ultimate efficiency of the code (i.e., to approach capacity, but not necessarily with an exponential behavior of the error probability) are not yet fully understood. The current results, based for example on the Shulman and Feder bound and its variants and some combinations with other bounds (see, e.g., [133, 176, 201, 200]) do provide partial answers (such as the adequacy of a sub-exponential deviation in terms of (4.39)), but definitely are not fully conclusive.

The generalized Gallager-based bound presented in Section 4.2.2.1 calls for a deeper understanding of the potential use of this generalized expression, and yet be able to provide insightful and compact closed form results. Namely, the ability to use efficiently the additional degrees of freedom provided by the code partition as well as the generalized functions/parameters, is not yet fully understood and calls for further study.

As commented by the authors in [204], there is room for further tightening lower bounds obtained with sphere-packing arguments for codes of small to moderate block lengths, especially when focusing on a particular channel. The improvements of the sphere-packing bounds, as introduced by Valembois and Fossorier, are valid for the general case and yield a bound of reasonable complexity in formulation and computation. These are applicable for assessing the theoretical limitations of block codes of moderate block lengths (e.g., turbo and LDPC codes). Aside from inspiring more research on this subject, the bounds introduced in [204] modify the 1967 sphere-packing bound of Shannon, Gallager and Berlekamp [184] in a way which makes them applicable to memoryless discrete-input and continuous output channels. Particularized for the binary-input Gaussian channel, the new bounds in [204] prove themselves (especially for high code rates) as a superior bounding technique as compared to the 1959 sphere-packing bound of Shannon [185] which was tailored for the Gaussian channel (but does not take into account the modulation). Sphere-packing lower bounds improve the understanding of the potential gain which can be achieved by optimal decoding.

Considering de Caen's based bounds and their recent improved versions [42, 39], an interesting open problem is the generalization of these bounds to ensembles of linear codes, so that these new versions are expressible in terms of the average distance spectrum of the ensemble.

The performance bounds reviewed here are based on basic features of codes and ensembles, namely their distance spectra. These bounds are readily applicable to assess ultimate performance of new classes of codes on graphs (e.g., ensembles of protograph LDPC codes whose distance spectra are obtained by the techniques introduced in [51, 72]).

The topics we chose to cover here within the most extensive area of performance analysis of optimal (ML) decoders, are definitely impacted by our subjective views. In this respect, we may have overlooked items which should be covered in more detail. The extensive reference list provided here is related to the central topics covered, and in this respect it can not be considered conclusive. Yet, we do trust that the references here along with the references therein do give a comprehensive picture of the relevant literature.

Acknowledgments

We are grateful to Idan Goldenberg, Gil Wiechman and Moshe Twitto for detecting typos and providing useful comments which improved the lucidity of the presentation. We are also grateful to the anonymous reviewers for their helpful and constructive comments. We wish to thank Asaf Cohen, Marc Fossorier and Rouheng Liu for authorizing us to use figures from their contributions for the tutorial. The work was supported by the EU 6th International Framework Programme via the NEWCOM Network of Excellence.

References

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

[2] E. Agrell, “Voronoi regions for binary linear block codes,” *IEEE Trans. on Information Theory*, vol. 42, pp. 310–316, January 1996.

[3] E. Agrell, “On the Voronoi neighbor ratio for binary linear block codes,” *IEEE Trans. on Information Theory*, vol. 44, pp. 3064–3072, November 1998.

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

[5] R. Annavajjala, A. Chockalingam, and L. B. Milstein, “Performance analysis of coded communication systems on Nakagami fading channels with selection combining diversity,” *IEEE Trans. on Communications*, vol. 52, pp. 1214–1220, July 2004.

[6] A. Ashikmin and A. Barg, “Minimal vectors in linear codes,” *IEEE Trans. on Information Theory*, vol. 44, pp. 2010–2017, September 1998.

[7] M. Aydinlik and M. Salehi, “Performance bounds for unequal error protecting turbo codes,” in *Proceedings 2005 Conference on Information Sciences and Systems (CISS 2005)*, (Johns Hopkins University, Maltimore, MD, USA), pp. 90–96, March 16–18, 2005.

[8] F. Babich, "On the performance of efficient coding techniques over fading channels," *IEEE Trans. on Wireless Communications*, vol. 3, pp. 290–299, January 2004.

[9] F. Babich, G. Montorsi, and F. Vatta, "Performance bounds of continuous and blockwise decoded turbo codes in Rician fading channel," *Electronics Letters*, vol. 34, pp. 1646–1648, August 1998.

[10] F. Babich, G. Montorsi, and F. Vatta, "Improved union bounds on turbo codes performance," *IEE Proceedings on Communications*, vol. 147, pp. 337–344, December 2000.

[11] F. Babich, G. Montorsi, and F. Vatta, "On the Viterbi and Viterbi's improved union bounds on turbo codes performance," in *Proceedings Second International Symposium on Turbo Codes and Related Topics*, (Brest, France), pp. 527–531, September 4–7, 2000.

[12] V. B. Balakirsky and A. J. H. Vinck, "Estimates of the decoding error probability for parallel channels with dependent noise," in *Proceedings IEEE International Symposium on Information Theory and its Applications (ISITA 2004)*, (Parma, Italy), pp. 1568–1573, October 2004.

[13] A. Barg, "On the asymptotic accuracy of the union bound," in *Proceedings Forty-Second Annual Allerton Conference on Communication, Control and Computing*, (Urbana-Champaign, IL, USA), pp. 1352–1361, September 2004.

[14] A. M. Barg and I. I. Dumer, "On computing the weight spectrum of cyclic codes," *IEEE Trans. on Information Theory*, vol. 38, pp. 1382–1386, July 1992.

[15] L. Bazzi, T. Richardson, and R. Urbanke, "Exact thresholds and optimal codes for the binary symmetric channel and Gallager's decoding algorithm A," *IEEE Trans. on Information Theory*, vol. 50, pp. 2010–2021, September 2004.

[16] F. Behnamfar, F. Alajaji, and T. Linder, "Improved lower bounds for the error rate of linear block codes," in *Proceedings 43rd Allerton Conference on Control, Computing and Communications*, (Monticello, Illinois, USA), pp. 2227–2236, September 2005.

[17] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Analysis, design, and iterative decoding of double serially concatenated codes with interleavers," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 231–244, February 1998.

[18] 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, pp. 909–926, May 1998.

[19] S. Benedetto and G. Montorsi, "Unveiling turbo codes: some results on parallel concatenated coding schemes," *IEEE Trans. on Information Theory*, vol. 42, pp. 409–429, March 1996.

[20] S. Benedetto and G. Montorsi, "Performance of continuous and blockwise decoded turbo codes," *IEEE Communications Letters*, vol. 1, pp. 77–79, May 1997.

[21] A. Bennatan and D. Burshtein, "On the application of LDPC codes to arbitrary discrete-memoryless channels," *IEEE Trans. on Information Theory*, vol. 50, pp. 417–438, March 2004.

[22] E. R. Berlekamp, “The technology of error correction codes,” *Proceedings of the IEEE*, vol. 68, pp. 564–593, May 1980.

[23] C. Berrou and A. Glavieux, “Near optimum error correcting coding and decoding: Turbo codes,” *IEEE Trans. on Communications*, vol. 44, pp. 1261–1271, October 1996.

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

[25] E. Biglieri, G. Caire, and G. Taricco, “Expurgating the union bound to error probability: A generalization of the Verdu-Shields theorem,” in *Proceedings of the 1997 IEEE International Symposium on Information Theory*, (Ulm, Germany), June 1997.

[26] E. Biglieri, J. Proakis, and S. Shamai, “Fading channels: Information-theoretic and communications aspects,” *IEEE Trans. on Information Theory*, vol. 44, pp. 2619–2692, October 1998.

[27] H. Bouzekri and S. L. Miller, “An upper bound on turbo codes performance over quasi-static fading channels,” *IEEE Communications Letters*, vol. 7, pp. 302–304, July 2003.

[28] M. Breiling and J. Huber, “A method for determining the distance profile of turbo codes,” in *Proceedings 3rd ITG Conference on Source and Channel Coding*, (Munich, Germany), pp. 219–224, January 2000.

[29] C. Brutel and J. Boutros, “Serial concatenation of interleaved convolutional codes and M-ary continuous phase modulations,” *Annals of Telecommunications*, vol. 54, pp. 235–242, March–April 1999.

[30] A. G. Burr, “Bounds on coding gain versus decoding delay for spherical codes on the Gaussian channel,” in *Proceedings 1997 IEEE International Symposium on Information Theory (ISIT 1997)*, (Ulm, Germany), June–July 1997.

[31] A. G. Burr, “Multilevel turbo-coded modulation: performance bounds,” in *Proceedings of the First International Symposium on Turbo Codes and Related Topics*, (Brest, France), pp. 111–118, September 1997.

[32] A. G. Burr and G. P. White, “Comparison of iterative decoder performance with union bounds for short frame turbo codes,” *Annals of Telecommunications*, vol. 54, pp. 201–207, March-April 1999.

[33] D. Burshtein, M. Krivelevich, S. Litsyn, and G. Miller, “Upper bounds on the rate of LDPC codes,” *IEEE Trans. on Information Theory*, vol. 48, pp. 2437–2449, September 2002.

[34] D. Burshtein and G. Miller, “Asymptotic enumeration methods for analyzing LDPC codes,” *IEEE Transactions on Information Theory*, vol. 50, pp. 1115–1131, June 2004.

[35] G. Caire and E. B. G. Taricco and, “Bit-interleaved coded modulation,” *IEEE Trans. on Information Theory*, vol. 44, pp. 927–946, May 1998.

[36] G. Caire, G. Taricco, and G. Batail, “Weight distribution and performance of the iterated product of single-parity check-codes,” in *Proceedings 1994 IEEE Global Telecommunications Conference (GLOBECOM '94), Communications Theory Mini-Conference*, pp. 206–211, December 1994.

[37] M. Cedervall and R. Johannesson, "A fast algorithm for computing the distance spectrum of convolutional codes," *IEEE Trans. on Information Theory*, vol. 35, pp. 1146–1159, November 1989.

[38] P. Chaudhari and A. K. Khandani, "Using the Fourier transform to compute the weight distribution of a binary linear block code," *IEEE Communications Letters*, vol. 5, pp. 22–24, January 2001.

[39] A. Cohen and N. Merhav, "Lower bounds on the error probability of block codes based on improvements on de Caen's inequality," *IEEE Trans. on Information Theory*, vol. 50, pp. 290–310, February 2004.

[40] J. Craig, "A new, simple and exact result for calculating error probability for two-dimensional signal constellation," in *Proceedings of 1991 Military Communications Conference (MILCOM '91)*, (Boston, MA, USA), pp. 25.5.1–25.5.5, November 1991.

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

[42] D. de Caen, "A lower bound on the probability of a union," *Discrete mathematics*, vol. 169, pp. 217–220, 1997.

[43] H. M. de Oliveira and G. Battail, "The random coded modulation: Performance and Euclidean distance spectrum evaluation," *Annals of Telecommunications*, vol. 47, pp. 107–124, March-April 1992.

[44] A. Dembo *personal communications*, communicated by I. Sason and S. Shamai, 2000.

[45] Y. Desaki, T. Fujiwara, and T. Kasami, "A method for computing the weight distribution of a block code by using its trellis diagram," *IEICE Trans. Fundamentals*, vol. E77-A, pp. 1230–1237, August 1994.

[46] Y. Desaki, T. Fujiwara, and T. Kasami, "The weight distributions of extended binary primitive BCH codes of length 128," *IEEE Trans. on Information Theory*, vol. 43, pp. 1364–1371, July 1997.

[47] C. Di, *Asymptotic and Finite-Length Analysis of Low-Density Parity-Check Codes.* Ph.D. dissertation, EPFL, Lausanne, Switzerland, September 2004.

[48] C. Di, A. Montanari, and R. Urbanke, "Weight distributions of LDPC code ensembles: combinatorics meets statistical physics," in *Proceedings 2004 IEEE International Symposium on Information Theory (ISIT 2004)*, (Chicago, USA), p. 102, July 2004.

[49] C. Di, T. Richardson, and R. Urbanke, "Weight distribution of low-density parity-check codes," accepted to *IEEE Trans. on Information Theory*, 2006.

[50] D. Divsalar, "A simple tight bound on error probability of block codes with application to turbo codes," the 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.

[51] D. Divsalar, "Ensemble weight enumerators for protograph LDPC codes," in *Proceedings IEEE 2006 International Symposium on Information Theory (ISIT 2006)*, (Seattle, Washington, USA), July 9–14 2006.

[52] D. Divsalar and E. Biglieri, "Upper bounds to error probabilities of coded systems beyond the cuttoff rate," *IEEE Trans. on Communications*, vol. 51, pp. 2011–2018, December 2003.

[53] D. Divsalar, S. Dolinar, R. J. McEliece, and F. Pollara, "Transfer function bounds on the performance of turbo codes," Jet Propulsion Laboratory (JPL), CA, USA, TDA Progress Report 42–122, August 1995. [Online] Available: http://tmo.jpl.nasa.gov/tmo/progressreport/42-122/122A.pdf.

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

[55] D. Divsalar and F. Pollara, "Serial and hybrid concatenated codes with applications," in *Proceedings of International Symposium on Turbo Codes and Related Topics*, pp. 80–87, September 1997.

[56] S. Dolinar and D. Divsalar, "Weight distributions for turbo codes using random and nonrandom permutations," Jet Propulsion Laboratory (JPL), CA, USA, TDA Progress Report 42–122, August 1995. [Online] Available: `http://tmo.jpl.nasa.gov/tmo/progress_report/42-122/122B.pdf`.

[57] S. Dolinar, D. Divsalar, and F. Pollara, "Code performance as a function of block size," Jet Propulsion Laboratory (JPL), TMO Progress Report 42–133, pp. 1–23, May 15, 1998. [Online] Available: `http://tmo.jpl.nasa.gov/tmo/progress_report/42-133/133K.pdf`.

[58] T. Duman and E. Kurtas, "Maximum likelihood decoding bounds for high rate turbo codes over Loretnzian channels," *Electronics Letters*, vol. 37, pp. 771–773, June 2001.

[59] T. Duman and M. Salehi, "The union bound for turbo coded modulation systems over fading channels," *IEEE Transactions on Communications*, vol. 47, pp. 1495–1502, (See also the correction in *IEEE Trans. on Communications*, vol. 50, p. 1766, November 1999), October 1999.

[60] 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.

[61] T. M. Duman and E. M. Kurtas, "Performance bounds for high rate linear codes over partial response channels," *IEEE Trans. on Information Theory*, vol. 47, pp. 1201–1205, March 2001.

[62] T. M. Duman and M. Salehi, "New performance bounds for turbo codes," *IEEE Trans. on Communications*, vol. 46, pp. 717–723, June 1998.

[63] T. M. Duman and M. Salehi, "Performance bounds for turbo-coded modulation systems," *IEEE Trans. on Communications*, vol. 47, pp. 511–521, April 1999.

[64] P. M. Ebert, *Error Bounds for Parallel Communication Channels*. Ph.D. dissertation, MIT, August 1966.

[65] M. El-Khamy and R. Garello, "On the weight enumerator and the maximum-likelihood performance of linear product codes," in *submitted to IEEE Trans. on Information Theory*, December 2005. [Online]. Available: http://arxiv.org/abs/cs.IT/0601095.

[66] M. El-Khamy and R. J. McEliece, "Bounds on the average binary minimum distance and the maximum-likelihood performance of Reed Solomon codes," in *Proceedings Forty-Second Annual Allerton Conference on Commu-*

nication, Control and Computing, (Urbana-Champaign, IL, USA), pp. 290–299, September 29–October 1, 2004.

[67] K. Engdahl and K. Zigangirov, "Tighter bounds on the error probability of fixed convolutional codes," *IEEE Trans. on Information Theory*, vol. 47, pp. 1625–1630, May 2001.

[68] U. Erez and G. Miller, "The ML decoding performance of LDPC ensembles over Z_q," *IEEE Trans. On Information Theory*, vol. 51, pp. 1871–1879, May 2005.

[69] T. Etzion, A. Trachtenberg, and A. Vardy, "Which codes have cycle-free Tanner graphs?," *IEEE Trans. on Information Theory*, vol. 45, pp. 2173–2181, September 1999.

[70] A. G. Fabregas and G. Caire, "Coded modulation in the block-fading channel: Coding theorems and code construction," *IEEE Trans. on Information Theory*, vol. 52, pp. 91–114, January 2006.

[71] R. M. Fano, *Transmission of Information.* Jointly published by the MIT Press and John Wiley & Sons, 1961.

[72] S. L. Fogal, R. McEliece, and J. Thorpe, "Enumerators for protograph ensembles of LDPC codes," in *Proceedings IEEE 2005 International Symposium on Information Theory (ISIT 2005)*, (Adelaide, Australia), pp. 2156–2160, September 4–9 2005.

[73] G. D. Forney and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Trans. on Information Theory*, vol. 44, pp. 2384–2415, October 1998.

[74] M. Fossorier, "Critical point for maximum-likelihood decoding of linear block codes," *IEEE Communications Letters*, vol. 9, pp. 817–819, September 2005.

[75] M. P. C. Fossorier, S. Lin, and D. J. Costello, "On the weight distribution of terminated convolutional codes," *IEEE Trans. on Information Theory*, vol. 45, pp. 1646–1648, July 1999.

[76] M. Fozunbal, S. W. McLaughlin, and R. W. Schafer, "On performance limits of space-time codes: A sphere-packing bound approach," *IEEE Trans. on Information Theory*, vol. 49, pp. 2681–2687, October 2003.

[77] S. Franz, M. Leone, A. Montanari, and F. R. Tersenghi, "Dynamic phase transition for decoding algorithms," *Physical Review E*, vol. 66, pp. 046120-1–046120-17, October 2002.

[78] F. Gagnon and D. Haccoun, "Bounds on the error performance of coding for non independent Rician Fading channels," *IEEE Trans. on Communications*, vol. 40, pp. 351–360, February 1992.

[79] J. Galambos and I. Simonelli, "Bonferroni-type inequalities with Applications," in *Springer Series in Statistics, Probability and its Applications*, (New-York), Springer-Verlag, 1996.

[80] R. G. Gallager, "Low-density parity-check codes," *IRE Trans. on Information Theory*, vol. 8, pp. 21–28, January 1962.

[81] R. G. Gallager, *Low-density parity-check codes.* Cambridge, MA, USA: MIT Press, 1963.

[82] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. on Information Theory*, vol. 11, pp. 3–18, January 1965.

[83] R. G. Gallager, *Information Theory and Reliable Communications.* 1968.

[84] R. G. Gallager, "The random coding bound is tight for the average code," *IEEE Trans. on Information Theory*, vol. 19, pp. 244–246, March 1973.

[85] A. Ganti, A. Lapidoth, and E. Telatar, "Mismatched decoding revisited: General alphabets, channels with memory, and the wide-band limit," *IEEE Trans. on Information Theory*, vol. 46, pp. 2315–2328, November 2000.

[86] I. S. Gradshteyn and I. M. Ryzhik, *Tables of integrals, series and products.* Academic Press, Fifth ed., 1994.

[87] S. Guemghar and G. Caire, "On the performance of finite length irregular repeat-accumulate codes," in *Proceedings of the Fifth ITG Conference on Source and Channel Coding*, (Germany), pp. 79–86, January 14–16, 2004.

[88] J. Ha, J. Kim, and S. W. McLaughlin, "Rate-compatible puncturing of low-density parity-check codes," *IEEE Trans. on Information Theory*, vol. 50, pp. 2824–2836, November 2004.

[89] E. K. Hall and S. G. Wilson, "Design and analysis of turbo codes on Rayleigh fading channels," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 160–174, February 1998.

[90] H. Herzberg and G. Poltyrev, "Techniques for bounding the probability of decoding error for block coded modulations structures," *IEEE Trans. on Information Theory*, vol. 40, pp. 903–911, May 1994.

[91] H. Herzberg and G. Poltyrev, "The error probability of M-ary PSK block coded modulation schemes," *IEEE Trans. on Communications*, vol. 44, pp. 427–433, April 1996.

[92] J. Hokfelt, O. Edfors, and T. Maseng, "On the theory and performance of trellis termination methods for turbo codes," *IEEE Journal on Selected Areas in Communications*, vol. 19, pp. 838–847, May 2001.

[93] C. H. Hsu and A. Anastasopoulos, "Asymptotic weight distributions of irregular repeat-accumulate codes," in *Proceedings 2005 IEEE Global Telecommunications Conference (GLOBECOM '05)*, (St Louis, Mo, USA), pp. 1147–1151, November 2005.

[94] C. H. Hsu and A. Anastasopoulos, "Capacity-achieving codes with bounded graphical complexity on noisy channels," in *presented in the 43rd Allerton Conference on Communication, Control and Computing*, (Monticello, Illinois, USA), pp. 28–30, September 2005. [Online]. Available: http://www.arxiv.org/abs/cs.IT/0509062.

[95] C. H. Hsu and A. Anastasopoulos, "Capacity-achieving LDPC codes through puncturing," in *Proceedings 2005 IEEE International Conference on Wireless Networks and Mobile Computing (WirelessCom 2005)*, (Mauii, Hawaii, USA), pp. 13–16, June 2005.

[96] J. Hu and S. L. Miller, "An improved upper bound on the performance of convolutional codes over quasi-static fading channels," in *Proceedings 2003 IEEE Global Communications Conference (GLOBECOM 2003)*, (San Francisco, CA, USA), pp. 1593–1597, December 1–5, 2003.

[97] X. L. Huang, N. Phamdo, and L. Ping, "BER bounds on parallel concatenated single parity-check arrays and Zigzag codes," in *Proceedings 1999 IEEE*

Conference on Global Communications (GLOBECOM '99), (Rio de Janeiro, Brazil), pp. 2436–2440, December 1999.

[98] B. Hughes, "On the error probability of signals in additive white Gaussian noise," *IEEE Trans. on Information Theory*, vol. 37, pp. 151–155, January 1991.

[99] L. W. Hughes, "A simple upper bound on the error probability for orthogonal signals in white noise," *IEEE Trans. on Communications*, vol. 40, p. 670, April 1992.

[100] D. Hunter, "An upper bound for the probability of a union," *Journal of Applied Probability*, vol. 13, pp. 597–603, 1976.

[101] E. A. Ince, N. S. Kambo, and S. A. Ali, "Efficient expression and bound for pairwise error probability in Rayleigh fading channels, with application to union bounds for turbo codes," *IEEE Communications Letters*, vol. 9, pp. 25–27, January 2005.

[102] J. Jiang and K. R. Narayanan, "Iterative soft input soft output decoding of Reed-Solomon codes by adapting the parity-check matrix," submitted to *IEEE Trans. on Information Theory*, [Online]. Available: http://www.arxiv.org/list/cs.IT/0506073, June 2005.

[103] H. Jin and R. J. McEliece, "RA codes achieve AWGN channel capacity, lecture notes in computer science," in *Proceedings in Applied Algebra, Algebraic Algorithms and Error-Correcting Codes: 13th International Symposium (AAECC-13)*, (M. Fossorier, H. Imai, S. Lin, and A. Poli, eds.), (Honolulu, Hawaii, USA), pp. 10–18, Springer-Verlag Heidelberg, November 1999.

[104] H. Jin and R. J. McEliece, "Coding theorems for turbo-like ensembles," *IEEE Trans. on Information Theory*, vol. 48, pp. 1451–1461, June 2002.

[105] G. Kaplan and S. Shamai, "Information rates and error exponents of compound channels with application to antipodal signaling in a fading environment," *International Journal of Electronics and Communication (AEU)*, vol. 47, pp. 228–239, April 1993.

[106] G. Kaplan and S. Shamai, "Achievable performance over the correlated Rician channel," *IEEE Trans. on Communications*, vol. 42, pp. 2967–2978, November 1994.

[107] T. Kasami, T. Fujiwara, and S. Lin, "Approximation of the weight distribution of binary linear block codes," *IEEE Trans. on Information Theory*, vol. 31, pp. 769–780, November 1985.

[108] O. Keren and S. Litsyn, "A lower bound on the probability of decoding error over a BSC channel," in *Proceedings 21st IEEE Convention of the Electrical and Electronic Engineers in Israel*, (Tel-Aviv, Israel), pp. 271–273, April 2000.

[109] R. Knopp and P. A. Humblet, "On coding for block fading channels," *IEEE Trans. on Information Theory*, vol. 46, pp. 189–205, January 2000.

[110] K. Koiko and H. Ogiwara, "Performance evaluation of turbo codes over impulsive noise channels," *IEICE Trans. Fundamentals*, vol. E84–A, pp. 2418–2426, October 2001.

[111] P. Komulainen and K. Pehkonen, "Performance evaluation of superorthognal turbo codes in AWGN and flat Rayleight fading channels," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 196–205, February 1998.

[112] T. Koumoto, "Web site on the weight distribution of BCH and Reed-Muller codes," [Online]. Available: http://www.infsys.cne.okayama-u.ac.jp/∼koumoto/wd/index.html., August 2004.

[113] H. Kuai, F. Alajaji, and G. Takahara, "A lower bound on the probability of a finite union of events," *Journal of Discrete Mathematics*, vol. 215, pp. 147–158, March 2000.

[114] A. Lapidoth and S. Shamai, "Fading channels: How perfect need "perfect side-information" be?," *IEEE Trans. on Information Theory*, vol. 48, pp. 1118–1130, May 2002.

[115] D. L. Lazic, T. Beth, and M. Calic, "How close are turbo codes to optimal codes?," in *Proceedings of the International Symposium on Turbo Codes and Related Topics*, (Brest, France), pp. 192–195, September 3–5, 1997.

[116] K. Leeuwin, J. C. Belfiore, and K. Kaleh, "Chernoff bound of trellis-coded modulation over correlated fading channels," *IEEE Trans. on Communications*, vol. 42, pp. 2506–2511, August 1994.

[117] J. Li, K. R. Narayanan, and C. N. Georghiades, "Generalized product-accumulate codes: Analysis and performance," in *Proceedings 2001 IEEE Global Communications Conference (GLOBECOM 2001)*, (San Antonio, Texas), pp. 975–979, November 25–29, 2001.

[118] J. Li, K. R. Narayanan, and C. N. Georghiades, "Product-accumulate codes: A class of codes with near-capacity performance and low decoding complexity," *IEEE Trans. on Information Theory*, vol. 50, pp. 31–46, January 2004.

[119] Y. Li and J. Moon, "Performance analysis of bit-interleaved space-time coding for OFDM in block fading channels," in *Proceedings 2004 IEEE Vehicular Technology Conference (VTC2004-Spring)*, (Milan, Italy), May 2004. [Online]. Available: http://www-cdslab.ece.umn.edu/library/papers/vtc04S.pdf.

[120] C. Ling and M. Fu, *personal communications.* October 2005.

[121] S. Litsyn and V. Shevelev, "On ensembles of low-density parity-check codes: Asymptotic distance distributions," *IEEE Trans. on Information Theory*, vol. 48, pp. 887–908, April 2002.

[122] 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, pp. 1405–1424, April 2006.

[123] H. Lu, P. V. Kumer, and E. Yang, "On the input-output weight enumerators of product accumulate codes," *IEEE Communications Letters*, vol. 8, pp. 520–522, August 2004.

[124] D. J. C. Mackay and R. M. Neal, "Near Shannon limit performance of low-density parity-check codes," *IEEE Electronic Letters*, vol. 33, pp. 457–458, March 1997.

[125] S. J. Macmullan and O. M. Collins, "A comparison of known codes, random codes and the best codes," *IEEE Trans. on Information Theory*, vol. 44, pp. 3009–3022, November 1998.

[126] E. Malkamaki and H. Leib, "Evaluating the performance of convolutional codes over block fading channels," *IEEE Trans. on Information Theory*, vol. 45, pp. 1643–1646, July 1999.

[127] R. J. McEliece, "How to compute weight enumerators for convolutional codes," in *Communications and Coding*, (M. Darnel and B. Honary, eds.), ch. 6, pp. 121–141, 1998.

[128] C. Measson, A. Montanari, T. Richardson, and R. Urbanke, "Life above threshold: From list decoding to area theorem and MSE," in *2004 IEEE Information Theory Workshop*, (San Antonio, TX, USA), October 24–29, 2004. [Online]. Available: http://www.arxiv.org/abs/cs.IT/0410028.

[129] C. Measson, A. Montanari, and R. Urbanke, "Maxwell construction: The hidden bridge between iterative and maximum a posteriori decoding," in *IEEE Trans. on Information Theory*, June 2005. submitted to [Online]. Available: http://www.arxiv.org/abs/cs.IT/0506083.

[130] C. Measson, A. Montanari, and R. Urbanke, "Why we can not surpass capacity: The matching condition," presented in the 43rd *Allerton Conference on Communication, Control and Computing*, pp. 28–30, [Online]. Available: http://www.arxiv.org/abs/cs.IT/0510045, September 2005.

[131] A. Mehrabian and S. Yousefi, "Improved tangential sphere bound on the ML decoding error probability of linear block codes in AWGN interference," in *Proceedings 2005 Conference on Information Sciences and Systems (CISS 2005)*, (Maltimore, MD, USA), pp. 176–181, Johns Hopkins University, March 16–18, 2005.

[132] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai, "On information rates for mismatched decoders," *IEEE Trans. on Information Theory*, vol. 40, pp. 1953–1967, November 1994.

[133] 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, pp. 2696–2710, November 2001.

[134] A. Mohammadi and W. Zhuang, "Variance of the turbo code performance bound over the interleavers," *IEEE Trans. on Information Theory*, vol. 48, pp. 2078–2086, July 2002.

[135] A. Montanari, "Turbo codes: The phase transition," *The European Physical Journal B*, vol. 18, pp. 121–136, November 2000.

[136] A. Montanari, "Tight bounds for LDPC codes and LDGM codes under MAP decoding," *IEEE Trans. on Information Theory*, vol. 51, pp. 3221–3246, September 2005.

[137] P. Moqvist and T. M. Aulin, "Serially concatenated continuous phase modulation with iterative decoding," *IEEE Trans. on Communications*, vol. 49, pp. 1901–1915, November 2001.

[138] A. H. Mugaibel and M. A. Kousa, "Evaluation of transfer functions for punctured turbo codes," *Electronics Letters*, vol. 36, pp. 805–807, April 2000.

[139] M. Namokel, "Error performance bounds of turbo codes employing nonuniform interleavers," in *Proceedings 1999 IEEE International Conference on Personal Wireless Communications*, (Jaipur, India), pp. 404–408, February 1999.

[140] K. R. Narayanan and G. L. Stüber, "A serial concatenation approach to iterative demodulation and decoding," *IEEE Trans. on Communications*, vol. 47, pp. 956–961, July 1999.

[141] K. R. Narayanan and G. L. Stüber, "Performance of trellis-coded CPM with iterative demodulation and decoding," *IEEE Trans. on Communications*, vol. 49, pp. 676–687, April 2001.

[142] N. Nefedov, "Evaluation of transfer functions for punctured turbo codes," in *Proceedings 2nd International Symposium on Turbo Codes & Related Topics*, (Brest, France), pp. 419–422, September 4–7, 2000.

[143] M. Oberg and P. H. Siegel, "Performance analysis of turbo-equalized partial response channels," *IEEE Trans. on Communications*, vol. 49, pp. 436–444, March 2001.

[144] T. Ohutsuki and J. M. Kahn, "BER performance of turbo-coded PPM CDMA systems on optical fiber," *IEEE Journal of Lightwave Technology*, vol. 18, pp. 1776–1784, December 2000.

[145] T. Ohutsuki and J. M. Kahn, "Transfer function bounds on performance of binary turbo coding followed by M-ary ortohognal signal mapping through interleaver," in *Proceedings 2000 IEEE International Conference on Communications (ICC 2000)*, (New Orleans, LA, USA), pp. 623–627, June 18–22, 2000.

[146] I. Onyszchuk, *Finding the complete path and weight enumerators of convolutional codes.* February 1990. TDA Progress Report 42–100, JPL.

[147] M. Peleg, I. Sason, S. Shamai, and A. Elia, "On interleaved, differntially encoded convolutional codes," *IEEE Trans. on Information Theory*, vol. 45, pp. 2572–2581, November 1999.

[148] H. D. Pfister, *On the capacity of finite state channels and the analysis of convolutional accumulate$-m$ codes.* March 2003.

[149] 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, pp. 2352–2379, July 2005.

[150] H. D. Pfister and P. H. Siegel, "The serial concatenation of rate–1 codes through uniform random interleavers," *IEEE Trans. on Information Theory*, vol. 49, pp. 1425–1438, June 2003.

[151] L. Ping, W. K. Leung, and K. Y. Wu, "Low-rate turbo-hadamard codes," *IEEE Trans. on Information Theory*, vol. 49, pp. 3213–3224, December 2003.

[152] G. Poltyrev, "Bounds on the decoding error probability of binary linear codes via their spectra," *IEEE Trans. on Information Theory*, vol. 40, pp. 1284–1292, July 1994.

[153] G. S. Poltyrev, "Random coding bounds for discrete memoryless channels," *Problems on Information Transmission*, vol. 18, pp. 9–21, translated from Problemy Predachi Informatsii (PPI), vol. 18, 12–26, 1982, January–March 1982.

[154] A. Ramesh, A. Chockalingam, and L. B. Milstein, "Bounds on the performance of turbo codes on Nakagami fading channels with diversity combining," in *Proceedings 2001 IEEE Global Telecommunications Conference (GLOBECOM 2001)*, (San Antonio, Texas, USA), pp. 1199–1204, November 25–29, 2001.

[155] C. T. Retter, "An average weight-distance enumerator for binary expansions of Reed-Solomon codes," *IEEE Trans. on Information Theory*, vol. 48, pp. 1195–1200, May 2002.

[156] T. Richardson, M. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. on Information Theory*, vol. 47, pp. 619–637, February 2001.

[157] T. Richardson and R. Urbanke, "Modern coding theory," in preparation. [Online]. Available: http://lthcwww.epfl.ch/papers/ics.ps.

[158] T. Richardson and R. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. on Information Theory*, vol. 47, pp. 599–618, February 2001.

[159] T. Richardson and R. Urbanke, "On the distribution of low weight codewords for turbo codes," in *Proceedings Forty-Second Annual Allerton Conference on Communication, Control and Computing*, (Urbana-Champaign, IL, USA), pp. 492–501, September 29–October 1, 2004.

[160] E. Rosnes and O. Ytrehus, "An efficient algorithm for tailbiting turbo code weight distribution calculation," in *Proceedings Third International Symposium on Turbo Codes and Related Topics*, (Brest, France), pp. 439–442, September 1–5, 2003.

[161] E. Rosnes and O. Ytrehus, "Improved algorithms for the determination of turbo code weight distributions," *IEEE Trans. on Communications*, vol. 53, pp. 20–26, January 2005.

[162] M. Rouanne and D. J. Costello, "An algorithm for computing the distance spectrum of trellis codes," *IEEE Journal on Selected Areas in Comunications*, vol. 7, pp. 929–940, August 1989.

[163] M. M. Salah, R. A. Raines, M. A. Temple, and T. G. Bailey, "Approach for deriving performance bounds of punctured turbo codes," *Electronics Letters*, vol. 35, pp. 2191–2192, December 1999.

[164] I. Sason, *Upper bounds on the maximum-likelihood decoding error probability for block codes and turbo-like codes.* Ph.D. dissertation, Technion–Israel Institute of Technology, Haifa, Israel, September 2001. [Online]. Available: http://www.ee.technion.ac.il/people/sason/phd.html.

[165] I. Sason and I. Goldenberg, "Coding for parallel channels: Gallager bounds and applications to repeat-accumulate codes," in *Proceedings of the 24th IEEE Convention of Electrical and Electronics Engineers in Israel*, (Eilat, Israel), pp. 334–338, November 15–17, 2006.

[166] I. Sason and I. Goldenberg, "Coding for parallel channels: Gallager bounds for binary linear codes with applications to turbo-like codes," *IEEE Trans. on Information Theory*, vol. 53, pp. 2394–2428, July 2007.

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

[168] I. Sason and S. Shamai, "Bounds on the error probability for block and turbo-block codes," *Annals of Telecommunications*, vol. 54, pp. 183–200, March–April 1999.

[169] I. Sason and S. Shamai, "Improved upper bounds on the ensemble performance of ML decoded low-density parity-check codes," *IEEE Communications Letters*, vol. 4, pp. 89–91, March 2000.

[170] I. Sason and S. Shamai, "Improved upper bounds on the ML decoding error probability of parallel and serial concatenated turbo codes via their ensemble

distance spectrum," *IEEE Trans. on Information Theory*, vol. 46, pp. 24–47, January 2000.

[171] I. Sason and S. Shamai, "On union bounds for random serially concatenated turbo codes with maximum likelihood decoding," *European Trans. on Telecommunications*, vol. 11, pp. 271–282, May–June 2000.

[172] 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, pp. 2275–2299, September 2001.

[173] I. Sason, S. Shamai, and D. Divsalar, "Tight exponential upper bounds on the ML decoding error probability of block codes over fully interleaved fading channels," *IEEE Trans. on Communications*, vol. 51, pp. 1296–1305, August 2003.

[174] 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, pp. 3052–3061, December 2002.

[175] I. Sason and R. Urbanke, "Information-theoretic lower bounds on the bit error probability of codes on graphs," in *Proceedings 2003 IEEE International Symposium on Information Theory*, (Yokohama, Japan), p. 268, June 29–July 4, 2003.

[176] I. Sason and R. Urbanke, "Parity-check density versus performance of binary linear block codes over memoryless symmetric channels," *IEEE Trans. on Information Theory*, vol. 49, pp. 1611–1635, July 2003.

[177] I. Sason and G. Wiechman, "Log-domain calculation of the SP59 sphere-packing bound with application to M-ary PSK block coded modulation," in *Proceedings IEEE Convention of Electrical and Electronics Engineers in Israel*, (Eilat, Israel), pp. 344–348, November 15–17, 2006.

[178] I. Sason and G. Wiechman, "On achievable rates and complexity of LDPC codes for parallel channels with application to puncturing," in *Proceedings IEEE 2006 IEEE International Symposium on Information Theory (ISIT 2006)*, (Seatle, Washington, USA), pp. 406–410, July 9–14, 2006.

[179] I. Sason and G. Wiechman, "Performance versus complexity per iteration for low-density parity-check codes: An information-theoretic approach," in *Proceedings of the Fourth International Symposium on Turbo Codes and Related Topics*, (Munich, Germany), April 3–7, 2006.

[180] 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, pp. 580–598, February 2007.

[181] C. Schlegel and L. Perez, "On error bounds and turbo codes," *IEEE Communications Letters*, vol. 3, pp. 205–207, July 1999.

[182] G. E. Seguin, "A lower bound on the error probability for signals in white Gaussian noise," *IEEE Trans. on Information Theory*, vol. 44, pp. 3168–3175, November 1998.

[183] 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.

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

[185] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell System Technical Journal*, vol. 38, pp. 611–656, May 1959.

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

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

[188] M. K. Simon and D. Divsalar, "Some new twists to problems involving the Gaussian probability integral," *IEEE Trans. on Communications*, vol. 46, pp. 200–210, February 1998.

[189] N. S. Skantzos, J. van Mourik, D. Saad, and Y. Kabashima, "Average and reliability error exponents in low-density parity-check codes," *Journal of Physics A*, vol. 36, pp. 11131–11141, October 2003.

[190] T. Sugita, T. Kasami, and T. Fujiwara, "The weight distribution of the third-order Reed-muller code of length 512," *IEEE Trans. on Information Theory*, vol. 42, pp. 1622–1625, September 1996.

[191] Y. V. Svirid, "Weight distributions and bounds for turbo codes," *European Transactions on Telecommunications*, vol. 6, pp. 543–555, September–October 1995.

[192] P. F. Swaszek, "A lower bound on the error probability for signals in white Gaussian noise," *IEEE Trans. on Information Theory*, vol. 41, pp. 837–841, May 1995.

[193] E. Telatar and R. Urbanke, "On the ensemble performance of turbo codes," in *Proceedings 1997 IEEE International Symposium on Information Theory (ISIT 97)*, (Ulm, Germany), p. 105, June 1997.

[194] C. Tellambura and A. D. S. Jayalath, "Generation of bivariate Rayleigh and Nakagami-m fading envelopes," *IEEE Communications Letters*, vol. 4, pp. 170–172, May 2000.

[195] L. Tolhuizen, "More results on the weight enumerator of product codes," *IEEE Trans. on Information Theory*, vol. 48, pp. 2573–2577, September 2002.

[196] L. Tolhuizen and C. Baggen, "On the weight enumerator of product codes," *Discrete Mathematics*, vol. 106–107, 483–488, 1992.

[197] L. Tolhuizen, C. Baggen, and E. H. Nowacka, "Union bounds on the performance of product codes," in *Proceedings 1998 IEEE International Symposium on Information Theory (ISIT 1998)*, (MIT, Cambridge, MA, USA), p. 267, August 15–21, 1998.

[198] H. M. Tullberg and P. H. Siegel, "Serial concatenated TCM with inner accumulate code – Part 1: Maximum-likelihood analysis," *IEEE Trans. on Communications*, vol. 53, pp. 64–73, January 2005.

[199] M. Twitto and I. Sason, "On the error exponents of some versions of improved tangential-sphere bounds," *IEEE Trans. on Information Theory*, vol. 53, pp. 1196–1210, March 2007.

[200] M. Twitto, I. Sason, and S. Shamai, "Tightened upper bounds on the ML decoding error probability of binary linear codes," 2006 IEEE International Symposium on Information Theory (ISIT 2006) Seatle, pp. 714–718, July 2006.

[201] M. Twitto, I. Sason, and S. Shamai, "Tightened upper bounds on the ML decoding error probability of binary linear block codes," to appear in the *IEEE Trans. on Information Theory*, vol. 53, April 2007.

[202] R. Urbanke, "LDPC asymptotic weight and stopping set spectrum calculator," [Online]. Available: http://lthcwww.epfl.ch/∼cdi/ldpc/ldpc_define.php.

[203] A. Valembois and M. Fossorier, "Box and match techniques applied to soft-decision decoding," *IEEE Trans. on Information Theory*, vol. 50, pp. 796–810, May 2004.

[204] A. Valembois and M. Fossorier, "Sphere-packing bounds revisited for moderate block length," *IEEE Trans. on Information Theory*, vol. 50, pp. 2998–3014, December 2004.

[205] S. Verdú, *Multiuser detection.* Cambridge University Press, 1998.

[206] A. J. Viterbi, ch. 8, pp. 242–244, *Principles of coherent communication.* McGraw-Hill, 1966.

[207] A. J. Viterbi and J. Omura, *Principles of digital communications and coding.* 1979.

[208] A. J. Viterbi, A. M. Viterbi, J. Nicolas, and N. T. Sindushayana, "Perspectives on interleaved concatenated codes with iterative soft-output decoding," in *Proceedings First International Symposium on Turbo Codes and Related Topics*, (Brest, France), pp. 47–54, September 3–5 1997.

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

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

[211] A. M. Viterbi and A. J. Viterbi, "New results on serial concatenated and accumulated convolutional turbo code performance," *Annals of Telecommunications*, vol. 54, pp. 173–182, March–April 1999.

[212] U. Wachsmann, R. F. H. Fischer, and J. B. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Trans. on Information Theory*, vol. 45, pp. 1361–1391, July 1999.

[213] T. Weijun, R. M. Todd, and J. R. Cruz, "Bounds for low-density parity-check codes over partial response channels," *IEEE Trans. on Magnetics*, vol. 38, pp. 2310–2312, September 2002.

[214] G. Wiechman and I. Sason, "Improved bounds on the parity-check density and achievable rates of LDPC codes," in *Proceedings Forty-Third Annual Aller-*

ton Conference on Communications, Control and Computing, pp. 1747–1758, September 28–30, 2005. See http://arxiv.org/abs/cs.IT/0505078.

[215] G. Wiechman and I. Sason, "Parity-check density versus performance of binary linear block codes over memoryless symmetric channels: new bounds and applications," *IEEE Trans. on Information Theory*, vol. 53, pp. 550–579, February 2007.

[216] J. K. Wolf and A. J. Viterbi, "On the weight distribution of linear block codes formed from convolutional codes," *IEEE Trans. on Communications*, vol. 44, pp. 1049–1051, September 1996.

[217] K. Wu, L. Ping, X. Huang, and N. Phamdo, "Performance analysis of turbo-SPC codes," *IEEE Trans. on Information Theory*, vol. 50, pp. 2490–2494, October 2004.

[218] X. Wu, Y. Cheng, and H. Xiang, "The Engdahl-Zigangirov bound for binary coded systems over block fading channels," *IEEE Communications Letters*, vol. 9, pp. 726–728, August 2005.

[219] X. Wu and H. Xiang, "New Gallager bounds in block fading channels," *IEEE Trans. on Information Theory*, vol. 53, pp. 684–694, February 2007.

[220] K. Yasunaga and T. Fujiwara, "Determination of the local weight distribution of binary linear block codes," *IEEE Trans. on Information Theory*, vol. 52, pp. 4444–4454, October 2006.

[221] T. Yokokawa, T. Miyauchi, K. Yamamoto, Y. Iida, M. Hattori, and R. J. McEliece, "ML Performance analysis method for SCCC and SCTCM with an 'in-line' interleaver," in *Proceedings Third International Symposium on Turbo Codes and Related Topics*, (Brest, France), pp. 303–306, September 1–5, 2003.

[222] H. Yoshikawa, "On the calculation method of input-output weight distribution of terminated convolutional codes," in *Proceedings 2003 International Symposium on Information Theory and Applications (ISITA 2004)*, (Parma, Italy), pp. 852–855, October 10–13, 2004.

[223] S. Yousefi and A. K. Khandani, "Generalized tangential-sphere bound on the ML decoding error probability of linear binary block codes in AWGN interference," *IEEE Trans. on Information Theory*, vol. 50, pp. 2810–2815, November 2004.

[224] S. Yousefi and A. K. Khandani, "A new upper bound on the ML decoding error probability of linear block codes in the AWGN channel," *IEEE Trans. on Information Theory*, vol. 50, pp. 3026–3036, December 2004.

[225] J. Zangl and R. Herzog, "Improved tangential sphere bound on the bit error probability of concatenated codes," *IEEE Journal on Selected Areas in Communications*, vol. 19, pp. 825–830, May 2001.

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

[227] S. A. Zummo and W. E. Stark, "Performance analysis of coded systems over block fading channels," in *Proceedings 2002 IEEE Vehicular Technology Conference (VTC 2002-Fall)*, pp. 1129–1133, September 24–28, 2002.

[228] S. A. Zummo and W. E. Stark, “Performance analysis of binary coded systems over rician block fading channels,” in *Proceedings 2003 IEEE Military Communications Conference (MILCOM 2003)*, pp. 314–319, October 13–16, 2003.