THEORETICAL AND PRACTICAL ASPECTS ON THE PERFORMANCE VERSUS COMPLEXITY TRADEOFF FOR LDPC-BASED CODES

GIL WIECHMAN

THEORETICAL AND PRACTICAL ASPECTS ON THE PERFORMANCE VERSUS COMPLEXITY TRADEOFF FOR LDPC-BASED CODES

RESEARCH THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

GIL WIECHMAN

SUBMITTED TO THE SENATE OF THE TECHNION — ISRAEL INSTITUTE OF TECHNOLOGY

Shebat 5768 HAIFA JANUARY 2008

THIS RESEARCH THESIS WAS SUPERVISED BY DR. IGAL SASON UNDER THE AUSPICES OF THE DEPARTMENT OF ELECTRICAL ENGINEERING

ACKNOWLEDGMENT

I wish to thank my supervisor, Dr. Igal Sason, for his dedicated guidance and for contributing greatly to the research and the analysis presented in this dissertation. The enjoyment and delight-in-doing which I have experienced during my time at the Technion is largely due to him. It has been a great privilege to work in a fruitful and enjoyable cooperation with a researcher of his caliber. I wish to thank him for instilling in me some of his curiosity and his strong desire to fully understand any problem we chose to tackle. Moreover, his warm and friendly attitude, patience, and true willingness to help and contribute in any professional and personal hardship will remain with me for years to come.

I also wish to thank my parents, Rivka and Menachem, for their encouragement and support. Ever since I can remember, education and scholarship have always been a first priority in my family. This work is undoubtedly a direct result of these values.

This work could not have been accomplished without the warm support of my wife, Anat. My heartfelt gratitude goes to her for her encouragement, support, and understanding during the numerous evening and nights when my mind had drifted away from the daily affairs of our home. She has been a perfect companion in the long and challenging journey which we traveled in the last few years.

The generous financial help of Andrew and Erna Finci Viterbi, the Ne'eman Foundation, and the Technion is gratefully acknowledged.

This research was supported by the Israel Science Foundation (grant no. 1070/07).

Contents

List of Figures

List of Tables

Abstract

Error-correcting codes which employ iterative decoding algorithms are now considered state of the art in the field of low-complexity coding techniques. The graphical representation of these codes is used to describe their algebraic structure, and also enables a unified description of their iterative decoding algorithms over various channels. These codes closely approach the capacity limit of many standard communication channels under iterative decoding. By now, there is a large collection of families of iteratively decoded codes including low-density parity-check (LDPC), low-density generator-matrix (LDGM), turbo, repeat-accumulate and their variants, zigzag, and product codes; all of them, demonstrate a rather small gap (in rate) to capacity with feasible complexity. The outstanding performance of these codes motivates an information-theoretic study of the tradeoff between their performance and complexity, as well as a study of the ultimate limitations of finite-length codes.

We begin our study of the performance versus complexity tradeoff by deriving bounds on the achievable rates and the graphical complexity of binary linear block codes under ML decoding. These bounds are derived under the assumption that the transmission takes place over memoryless binary-input output-symmetric (MBIOS) channels. The bounds are particularized to LDPC codes, and apply to the tradeoff between achievable rates and decoding complexity per iteration under message-passing decoding. Further, we generalize these bounds for the case where the codes are transmitted over a set of independent parallel MBIOS channels. The latter results are applied to ensembles of punctured LDPC codes.

Secondly, we consider the number of iterations required for successful iterative message-passing decoding of graph-based codes. The communication (this time) is assumed to take place over the binary erasure channel, and the analysis refers to the asymptotic case where the block length tends to infinity. We derive rigorous lower bounds on the number of decoding iterations required to achieve a given bit erasure probability under standard iterative message-passing decoding. These bounds are expressed in terms of the desired bit erasure probability and the gap between the design rate of the ensemble and the channel capacity. Ensembles of LDPC codes and the

more recently introduced families of systematic and non-systematic irregular repeat-accumulate and systematic accumulate-repeat-accumulate codes are considered. For all these code families, we show that the number of iterations scales at least like the inverse of the multiplicative gap to capacity; this matches a previous conjecture and experimental results.

Finally, we consider sphere-packing lower bounds on the decoding error probability of optimal block codes. We focus on modifications to the 1967 sphere-packing (SP67) bounding technique to make it more attractive for codes of finite block lengths. We derive a new sphere-packing bound (called the ISP bound) targeted at finite-length block codes transmitted over symmetric memoryless channels. This part of the work facilitates the assessment of the fundamental limitations of finite-length block codes, and is therefore very applicative for the evaluation of practical coded communication systems.

Notation

- x – Scalar.
- $\mathbf{x}$ – Row vector.
- X – Matrix.
- $\mathcal{X}$ – Set.
- x_i – The i'th element of the vector $\mathbf{x}$.
- $|\cdot|$ – Absolute value.
- $||\cdot||$ – Standard Euclidian norm.
- $\cdot^T$ – Tranpose.
- $\mathbb{E}(X)$ – Expectation of X.
- $H(X)$ – Entropy of X in bits.
- $H(X|Y)$ – Conditional entropy of X given Y, in bits.
- $h_2(\cdot)$ – Binary entropy function in base 2.
- $f'(x)$ – the first derivative of the function f with respect to x.
- $f''(x)$ – the second derivative of the function f with respect to x.
- $f(\varepsilon) = O\big(g(\varepsilon)\big)$ – there exist positive constants c and δ, such that $0 \leq f(\varepsilon) \leq cg(\varepsilon)$ for all $0 \leq \varepsilon \leq \delta$.
- $f(\varepsilon) = \Omega\big(g(\varepsilon)\big)$ – there exist positive constants c and δ, such that $0 \leq cg(\varepsilon) \leq f(\varepsilon)$ for all $0 \leq \varepsilon \leq \delta$.
- $f(\varepsilon) = \Theta\big(g(\varepsilon)\big)$ – there exist positive constants c_1, c_2 and δ, such that $0 \leq c_1 g(\varepsilon) \leq f(\varepsilon) \leq c_2 g(\varepsilon)$ for all $0 \leq \varepsilon \leq \delta$.

Abbreviations

- ARA – Accumulate-repeat-accumulate
- AWGN – Additive white Gaussian noise
- BEC – Binary erasure channel
- BPSK – Binary phase shift keying
- BSC – Binary symmetric channel
- CLB – Capacity limit bound
- dB – DeciBell
- d.d. – Degree distribution
- DE – Density-evolution
- DGLDPC – Doubly generalized low-density parity-check
- EXIT – Extrinsic information transfer
- GEXIT – Generalized extrinsic information transfer
- GLDPC – Generalized low-density parity-check
- i.i.d. – Independent identically distributed
- IP-LDPC – Intentionally punctured low-density parity-check
- IRA – Irregular repeat-accumulate
- ISP – Improved sphere-packing
- LDGM – Low-density generator-matrix
- LDPC – Low-density parity-check

- LHS – Left hand side
- LLR – Log-likelihood ratio
- MAP – Maximum a-posteriori
- MBIOS – Memoryless binary-input output-symmetric
- ML – Maximum likelihood
- PSK – Phase shift keying
- QPSK – Quadrature phase shift keying
- RC – Random coding
- RHS – Right hand side
- RP-LDPC – Randomly punctured low-density parity-check
- SP59 – The 1959 sphere-packing bound
- SP67 – The 1967 sphere-packing bound
- SPC – Single parity-check.
- TSB – Tangential-sphere bound
- VF – Valembois Fossorier

Chapter 1

Introduction

The mathematical foundations of information theory were laid by Shannon in 1948 [88]. One of the most surprising results introduced in this groundbreaking work is that information can be communicated with arbitrarily small distortion at positive rates, the highest of which is known as the channel capacity. Shannon's solution to the communication problem relies on using random block codes. The random nature of these codes implies that memory requirements of the encoder, as well as the memory and time requirements of the decoder, grow exponentially with the block length. Therefore, while the random codes serve as a fundamental tool in an innovative existence proof, they are of little practical use. This elementary result in information theory led to the birth of coding theory whose aim is to design *practical* coding and decoding schemes which approach the fundamental limitations set by Shannon. We refer the reader to a recent survey paper by Costello and Forney [20] which traces the evolution of efficient coding schemes since the landmark paper of Shannon. In the following, we briefly describe families of error-correcting codes which are addressed in this dissertation.

1.1 Linear Block Codes

Much of the effort of coding theorists focuses on *linear* block codes. These codes facilitate a substantial reduction in the space requirement, while still maintaining the potential of achieving reliable communications at rates arbitrarily close to the Shannon capacity limit (see [32, Section 6.2], [110, Section 3.10]). Linear block codes can be represented by a *generator matrix* whose rows form basis vectors of the code space. Alternatively, the code may be represented by a *parity-check matrix* whose rows form a basis of the vector space which is orthogonal to the code. Both of these approaches reduce the memory requirements for storing the code to the order of the

squared block length. The algebraic structure of linear block codes also enables the application of certain shortcuts in the decoding process, making it more computationally feasible. Notable early examples of linear block codes include the well-known codes of Hamming and Golay. Another prominent family of linear block codes are the Reed-Solomon (RS) codes and the related Bose-Chaudhuri-Hocquenghem (BCH), generalized RS, and alternant codes. Together with their elegant decoding algorithms, these codes are used in a wide range of common applications. Further details on algebraic coding schemes can be found in [48, 75] and references therein.

A common approach to the design of linear block codes focuses on enlarging the minimal distance of the codes, i.e., the Hamming distance between the two closest codewords. Researchers following this approach employ sophisticated algebraic tools to construct codes with large minimum distances, whose structure enables to increase the maximal number of channel errors which can be corrected with certainty. However, codes constructed using this approach fail to achieve capacity-approaching performance on many important communication channel models. Nevertheless, it should be noted that recently new algorithms which allow list decoding of such algebraic codes (see e.g., [75, Chapter 9], [33]) demonstrate a remarkable improvement in the performance of these codes over a variety of communication channels.

1.2 Gallager's LDPC Codes

Low-density parity-check (LDPC) codes form a subclass of linear block codes. In general, LDPC codes are linear block codes which can be represented by sparse parity-check matrices. These codes, along with the concept of their efficient decoding algorithms, were introduced by Gallager in his 1961 Ph.D. dissertation [30]. In Gallager's construction, the number of non-zero entries in every row of the parity-check matrix is fixed and the same property holds for the columns of this matrix. An (n, j, k) LDPC code is defined as a binary linear block code of length n, which is represented by parity-check matrix containing exactly j ones in each column and k ones in each row. Gallager provided the following procedure to construct these codes: First, divide the columns of the parity-check matrix H into j equal-size sections, creating j submatrices. Each of these submatrices will contain a single '$1'$ entry in each column. The first submatrix is constructed so that the i'th row contains 1's in columns $(i-1)k+1$ to ik, creating a sort of '*staircase*'. All the other submatrices are created by column permutations of the first submatrix. Note that in this construction, the number of ones in each row, k, must divide the block length n. Since each of the j submatrices is composed of exactly n/k rows, the rate of the code satisfies $R \geq 1 - j/k$. Gallager

defined the ensemble of (n, j, k) LDPC codes as the set of all codes constructed as above, using all possible column permutations.

Consider a linear block code and refer to each code symbol as a variable. A codeword of the linear code corresponds to an assignment of values to the code variables which satisfies a set of linear constraints defined by the rows of the parity-check matrix. ML decoding of a linear block code therefore amounts to finding the most likely assignment based on the channel input which satisfies these constraints. Note that for the binary codes considered by Gallager, these linear constraints amount to parity constraints on different subsets of the code bits.

One of the main novelties of [30] is the concept of applying efficient iterative decoding algorithms whose complexity scales linearly with the block length. The principle behind these algorithms is to treat each parity constraint and each variable localy. In the first part of each iteration of the algorithm, each code variable would exploit its corresponding channel input, as well as the information it received from the parity constraints it is involved in, to produce a probability assignment for its possible values. In the second part of the iteration, each constraint utilizes the probability assignments received from its participating variables to produce its own estimate of the probabilities for the possible values of each participating variable. In order to abstain from 'self persuasion', the variables do not take into account the message from a parity constraint when producing the message to the same constraint in the next iteration. Similarly, a message from a constraint to each participating variable is based only on information provided by the other variables involved in this constraint. The solution provided by the above local algorithm is clearly suboptimal. However, due to the sparseness of the parity-check matrices, the algorithms yield high performance while maintaining low complexity. Gallager suggested several iterative decoding algorithms based on the above approach which differ in the way that messages from variables to constraints and vice versa are calculated. Two of these algorithms apply to the binary symmetric channel (BSC) and a third applies to general memoryless binary-input output-symmetric channels (MBIOS).

1.3 The Re-Discovery of Graph-Based Codes

For more than three decades, LDPC codes and their iterative decoding algorithms were largely ignored by the coding theory community. One of the few notable exceptions is the work by Tanner [100] which introduces the notion of representing LDPC codes using graphs. The iterative decoding algorithms could now be understood as passing messages between variable nodes and check nodes over the edges of the graph.

The revival of graph-based codes and their iterative decoding algorithms in the mid 1990's is largely due to the phenomenal performance of turbo codes under practical iterative decoding algorithms [15]. This breakthrough triggered the re-discovery and generalization of LDPC codes [54], and the subsequent introduction of various other families of graph-based codes. By now, there is a large collection of families of graph-based codes, including LDPC, turbo, low-density generator-matrix (LDGM) [53], repeat-accumulate (RA) and their variants [24, 40, 4], product [28], and many other code families. All of them demonstrate excellent performance under practical iterative decoding algorithms.

The developments in the construction of graph-based codes has also led to a growing interest in the analytical study of the performance of efficient iterative decoding algorithms associated with these codes. When the graph representing the code does not contain cycles, then the iterative algorithm of Gallager for MBIOS channels [30] (known as the belief-propagation algorithm) is actually a bitwise maximum a-posteriori (MAP) decoding algorithm (See [74, Sections 2.5.1, 2.52]). A similar message-passing algorithm with different message update rules was shown to be an efficient blockwise MAP decoder for cycle-free codes [74, Section 2.5.5]. Unfortunately, it was also shown that cycle-free codes perform poorly even under MAP decoding [103]. A prominent development in the understanding of the performance of graph-based codes under iterative decoding algorithms was provided in [73], with the introduction of the density-evolution (DE) technique. The core concept behind this technique is to treat the messages passed along the edges of the graph during the iterative decoding process as random variables, and track the evolution of their probability density functions through the iterative process. DE analyzes the average performance over an ensemble of codes which are transmitted over an arbitrary MBIOS channel, and it applies to the asymptotic case where the block length tends to infinity. The analysis hinges on the fact that with probability 1, as the block length tends to infinity, the messages passed along different edges during any finite iteration are statistically independent from each other. This is known as the *tree assumption*, and it is the key factor which makes the analysis of the message densities feasible in the asymptotic case where we let the block length of these codes tend to infinity. It was also shown in [73] that the performance of individual codes concentrates around the average ensemble performance as the block length tends to infinity.

For general MBIOS channels, DE involves an infinite-dimensional analysis. The only exception to this is the BEC, where DE is simplified to a one-dimensional analysis. The extrinsic information transfer (EXIT) charts, pioneered by Stephan ten Brink [101, 102], form a powerful tool for an efficient design of codes defined on graphs by

tracing the convergence behavior of their iterative decoders. EXIT charts provide a good approximative engineering tool for tracing the convergence behavior of soft-input soft-output iterative decoders; they suggest a simplified visualization of the convergence of these decoding algorithms, based on a single parameter which represents the exchange of extrinsic information between the constituent decoders. For the BEC, the EXIT charts coincide with the DE analysis (see [74]). More recently, generalized extrinsic information transfer (GEXIT) charts were introduced [59]. These tools simplify the analysis of the iterative decoding process in the asymptotic case where the block length tends to infinity to a one-dimensional problem. Using GEXIT charts, links between belief propagation and MAP decoding have been exposed [60]. The development of these techniques relies on notions originally developed in statistical physics.

Due to the simplicity of the DE analysis for the BEC, proper design of codes defined on graphs enables to asymptotically achieve the capacity of the BEC under iterative message-passing decoding. Capacity-achieving sequences of LDPC ensembles were originally introduced by Shokrollahi [94] and Luby et al. [51], and a systematic study of capacity-achieving sequences of LDPC ensembles was presented by Oswald and Shokrollahi [63] for the BEC. Suitable constructions of capacity-achieving ensembles of variants of RA codes were devised in [40], [64], [65] and [80]. All these works rely on the DE analysis of codes defined on graphs for the BEC, and provide an asymptotic analysis which refers to the case where one lets the block length of these code ensembles tend to infinity. Another innovative coding technique, introduced by Shokrollahi [95], enables to achieve the capacity of the BEC with encoding and decoding complexities which scale linearly with the block length, and it has the additional pleasing property of achieving the capacity without the knowledge of the erasure probability of the channel. EXIT charts and Gaussian approximation of the message densities [18] have facilitated the design of graph-based codes which perform extremely well on a variety of common communication channels (See e.g., [5, 19, 23, 25, 26, 27, 37, 91] and references therein). The success of these codes has led to an increasing use of graph-based codes in a wide range of common applications [2, 3, 14].

The performance analysis of finite-length LDPC code ensembles whose transmission takes place over the BEC was introduced by Di et al. [22]. This analysis considers sub-optimal iterative message-passing decoding as well as optimal maximum-likelihood decoding. In [6], an efficient approach to the design of LDPC codes of finite length was introduced by Amraoui et al.; this approach is specialized for the

BEC, and it enables to design such code ensembles which perform well under iterative decoding with a practical constraint on the block length. In [72], Richardson and Urbanke initiated the analysis of the distribution of the number of iterations needed for the decoding of LDPC ensembles of finite block length which are communicated over the BEC. For general MBIOS channels, rigorous finite-length analysis of the performance of graph-based codes under iterative decoding algorithms is still in its infancy. A comprehensive reference on the construction and analysis of graph-based codes is given in [74].

1.4 Motivation and Related Work

In this work, we investigate the information-theoretic limitations on the performance versus complexity tradeoff of graph-based codes. The research is highly motivated by the outstanding performance of these codes over a variety of communication channels, while still preserving practical encoding and decoding complexity. The exceptional performance of graph-based codes with short to moderate block lengths also motivates a theoretical study of the performance limitations of finite-length block codes. The research is driven by the following core questions:

1. How good can LDPC codes be, even under optimal decoding?

2. What are the fundamental limitations on the complexity of iterative decoding algorithms, as a function of the gap between the code rate and the channel capacity?

3. What are the fundamental limitations on the performance of finite-length block codes?

We follow an innovative approach for characterizing the complexity of iterative decoders suggested by Khandekar and McEliece (see [42, 43, 56]). Their questions and conjectures were related to the tradeoff between the asymptotically achievable rates and the complexity under iterative message-passing decoding; they initiated a study of the encoding and decoding complexity of graph-based codes in terms of the achievable gap (in rate) to capacity. They conjectured that for a large class of channels, if the design rate of a suitably designed ensemble forms a fraction $1 - \varepsilon$ of the channel capacity, then the decoding complexity scales like $\frac{1}{\varepsilon} \ln \frac{1}{\varepsilon}$. The logarithmic term in this expression was attributed to the decoding complexity per iteration, and the number of iterations was conjectured to scale like $\frac{1}{\varepsilon}$. There is one exception: For the BEC, the complexity under the iterative message-passing decoding algorithm

behaves like $\ln \frac{1}{\varepsilon}$ (see [51], [80], [81] and [94]). This is true since the absolute reliability provided by the BEC allows every edge in the graph to be used only once during the iterative decoding. Hence, for the BEC, the number of iterations performed by the decoder serves mainly to measure the delay in the decoding process, while the decoding complexity is closely related to the complexity of the Tanner graph which is chosen to represent the code.

In his thesis [30], Gallager proved that right-regular LDPC codes (i.e., LDPC codes with a constant degree (a_{R}) of the parity-check nodes) cannot achieve the channel capacity on a BSC, even under ML decoding. This inherent gap to capacity is well approximated by an expression which decreases to zero exponentially fast in a_{R}. Richardson et al. [71] have extended this result, and proved that the same conclusion holds if a_{R} designates the *maximal right degree* of an irregular ensemble. Sason and Urbanke later observed in [81] that the result still holds when considering the *average right degree*. Gallager's bound [30, Theorem 3.3] provides an upper bound on the rate of right-regular LDPC codes which achieve reliable communications over the BSC. Burshtein et al. have generalized Gallager's bound for general ensembles of LDPC codes transmitted over general MBIOS channels [17]; to this end, they relied on a two-level quantization to the log-likelihood ratio (LLR) of these channels which essentially equates the available channel information to that of a physically degraded BSC.

Consider the number of ones in a parity-check matrix which represents a binary linear block code, and normalize it per information bit (i.e., with respect to the dimension of the code). This quantity (which is defined as the *density* of the parity-check matrix) is equal to the normalized number of left to right (or right to left) messages per information bit which are passed in the corresponding bipartite graph during a single iteration of the message-passing decoder. In [81], Sason and Urbanke considered the sparseness of parity-check matrices of binary linear block codes as a function of their gap to capacity (where, in general, this gap depends on the channel and on the decoding algorithm). An information-theoretic lower bound on the asymptotic density of parity-check matrices was derived in [81, Theorem 2.1] where this bound applies to every MBIOS channel and *every* sequence of binary linear block codes achieving a fraction $1 - \varepsilon$ of the channel capacity with vanishing bit error probability. It holds for an arbitrary representation of these codes by full-rank parity-check matrices, and is of the form $\frac{K_1 + K_2 \ln \frac{1}{\varepsilon}}{1-\varepsilon}$ where K_1 and K_2 are constants which only depend on the channel. Though the logarithmic behavior of this lower bound is in essence correct (due to a logarithmic behavior of the upper bound on the asymptotic parity-check density in [81, Theorem 2.2]), the lower bound in [81, Theorem 2.1] is

not tight (with the exception of the BEC, as demonstrated in [81, Theorem 2.3], and possibly also the BSC).

In numerous modern applications (see e.g. [2, 3]) the code rate may vary according to the level of noise currently present in the communication channel. The standard technique to achieve these varying transmission rates relies on using one base code and puncturing the encoder output according to a puncturing pattern associated with the desired rate. The main advantage of this method is that the same encoder and decoder are used, regardless of the communication rate. The performance of punctured LDPC codes under ML decoding was studied in [39] via analyzing the asymptotic growth rate of their average weight distributions and using upper bounds on the decoding error probability under ML decoding. Based on this analysis, it was proved that for any MBIOS channel, capacity-achieving codes of any desired rate can be constructed by puncturing the code bits of ensembles of LDPC codes whose design rate (before puncturing) is sufficiently low. The performance of punctured LDPC codes over the AWGN channel was studied in [35] under iterative message-passing decoding. Ha and McLaughlin studied in [35] two methods for puncturing LDPC codes where the first method assumes random puncturing of the code bits at a fixed rate, and the second method assumes possibly different puncturing rates for each subset of code bits which corresponds to variable nodes of a fixed degree. For the second approach, called 'intentional puncturing', the degree distributions of the puncturing patterns were optimized in [34, 35] where it was aimed to minimize the threshold under iterative decoding for a given design rate via the Gaussian approximation. Exact values of these optimized puncturing patterns were also calculated by the DE analysis, and they show good agreement with results obtained by the Gaussian approximation. The results in [34, 35] exemplify the usefulness of punctured LDPC codes for a relatively wide range of rates, and therefore, they are suitable for rate-compatible coding.

The bounds in [80, 81] and Chapters 2 and 3 of this dissertation show that the graphical complexity of capacity-approaching ensembles of un-punctured LDPC and systematic irregular repeat-accumulate (IRA) codes tends to infinity as the gap to capacity vanishes. We note that this result is mainly due to the relatively simple graphical structure of these codes. An additional degree of freedom which is obtained by introducing state nodes in the graph (e.g., punctured bits) was exploited in [64] and [65] to construct capacity-achieving ensembles of graph-based codes for the BEC which achieve an improved tradeoff between complexity and achievable rates. Surprisingly, these capacity-achieving ensembles under iterative decoding were demonstrated to maintain a *bounded complexity* per information bit regardless of the erasure probability of the BEC. A similar result of bounded complexity for capacity-achieving

ensembles over the BEC was also obtained in [38].

Sphere-packing bounds are lower bounds on the decoding error probability of optimal block codes. These bounds are expressed in terms of the block length, code rate, and communication channel. The 1959 sphere-packing (SP59) bound of Shannon [89] serves for the evaluation of the performance limits of block codes whose transmission takes place over an AWGN channel. This bound does not take into account the modulation used, but only assumes that the signals are of equal energy. It is often used as a reference for quantifying the sub-optimality of error-correcting codes under some practical decoding algorithms. The 1967 sphere-packing (SP67) bound, derived by Shannon, Gallager and Berlekamp [87], applies to optimal block codes transmitted over arbitrary discrete memoryless channels. Like the random coding bound of Gallager [31], the SP67 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 tight at the portion of the rate region between the critical rate (R_{c}) and the channel capacity; for all rates in this range, the error exponents of the SP67 and the random coding bounds coincide (see [87, Part 1]). In spite of its exponential behavior, the SP67 bound appears to be loose for codes of small to moderate block lengths. This weakness is due to the original focus in [87] on asymptotic analysis. In [109], Valembois and Fossorier revisit the derivation of the SP67 bound in order to improve its tightness for finite-length block codes (especially, for codes of short to moderate block lengths). As a side-effect of this improvement, the validity of the bound in [109] is extended to memoryless continuous-output channels (e.g., the binary-input AWGN channel). The remarkable improvement of their bound over the classical SP67 bound was exemplified in [109]; moreover, it provides an interesting alternative to the SP59 bound which is particularized for the AWGN channel [89].

1.5 This Dissertation

In Chapter 2, we introduce upper bounds on the achievable rates of binary linear block codes under ML decoding. We also derive lower bounds on the asymptotic density of an arbitrary presentation of their parity-check matrices as a function of the achievable gap (in rate) to capacity. These bounds are derived under the assumption that the transmission takes place over an MBIOS channel, and they refer to the case where the block length of the codes tends to infinity. The derivation of the bounds is motivated by the desire to improve previously reported bounds (see [17, Theorems 1 and 2] and [81, Theorem 2.1]) whose derivation relies on a two-level quantization of the LLR which therefore takes partial advantage of the available channel information.

An analysis based on a two-level quantization of the LLR, which in essence relies on information available from a physically degraded BSC in place of the actual MBIOS channel, is first modified to an analysis based on information from a quantized channel which better reflects the statistics of the actual communication channel (though the quantized information is still degraded w.r.t. the original information provided by the channel). The number of quantization levels of the LLR for the new channel used in the analysis is set to an arbitrary integer power of 2, and the calculation of these bounds is subject to an optimization of these quantization levels, as to obtain the tightest bounds within their form. The analysis is then modified to rely on the conditional pdf of the LLR at the output of the original MBIOS channel, and utilizes information available from an equivalent channel (without degrading the channel information). This second approach clearly leads to bounds which are uniformly tighter than the bounds derived via analysis of quantized channel information, and are surprisingly also easier to calculate. The significance of the bounds, using both quantized and un-quantized information, stems from a comparison between these bounds; such a comparison gives some insight on the effect of the number of quantization levels of the LLR (even if they are optimally determined) on the achievable rates, as compared to the ideal case where no quantization is performed. Chapter 2 is a reprint of the journal paper [118] (some of these results were also published in the conference papers [116, 84]).

In Chapter 3, we generalize the bounds introduced in Chapter 2 to the scenario where the codes are transmitted over the set of statistically independent parallel MBIOS channels. In this setup, each code bit is assigned to one specific communication channel. The transmission of punctured codes over a single channel can be regarded as a special case of communication of the original code over a set of parallel channels (which are defined by the puncturing rates applied to subsets of the code bits). We therefore apply the bounds on the achievable rates and decoding complexity of LDPC codes over parallel channels to the case of transmission of ensembles of punctured LDPC codes over an MBIOS channel. We state puncturing theorems related to achievable rates and decoding complexity of punctured LDPC codes. For ensembles of punctured LDPC codes, the calculation of bounds on their thresholds under ML decoding and their exact thresholds under iterative decoding (based on DE analysis) is of interest in the sense that it enables one to distinguish between the loss due to iterative decoding and the loss due to the structure of the ensembles. This chapter concludes with a diagram which shows interconnections between the theorems introduced in Chapters 2,3 and some other previously reported results [17, 65, 69, 67, 81]. Chapter 3 is a reprint of the journal paper [85] (the results are

also presented in part in the conference papers [83, 84]).

In Chapter 4, the number of iterations which is required for successful message-passing decoding of some important families of graph-based code ensembles (including LDPC and variations of RA codes) is considered. We present lower bounds on the number of decoding iterations for the case where the transmission of the code ensembles takes place over a BEC. These bounds refer to the asymptotic case where we let the block length tend to infinity. The bounds derived in this chapter are easily evaluated and are expressed in terms of some basic parameters of the ensemble which include the fraction of degree-2 variable nodes, the target bit erasure probability and the gap between the channel capacity and the design rate of the ensemble. It is demonstrated that the number of iterations which is required for successful message-passing decoding scales at least like the inverse of the gap to capacity, provided that the fraction of degree-2 variable nodes of the ensembles does not vanish (this condition is shown to hold for capacity-achieving LDPC ensembles under mild requirements, as shown in [76]). This asymptotic scaling of the lower bound on the number of iterations holds for various families of turbo-like code ensembles. Note that this is in contrast to the limitations on the graphical complexity, which scales differently for these different code families (see [64, 65, 80, 81]). The behavior of the lower bounds derived in Chapter 4 matches well with the experimental results and the conjectures on the number of iterations and complexity, as provided by Khandekar and McEliece (see [43, 42, 56]). The analysis in Chapter 4 relies on EXIT charts and on the area theorem for the BEC [9]. The analysis of the number of iterations for variations of RA codes also relies on the 'graph reduction' technique introduced in [64, Section II.C.2]. This chapter is a preprint of [86].

In Chapter 5 we focus on sphere-packing lower bounds on the decoding error probability of optimal block codes. We derive an improved sphere-packing bound (referred to as the 'ISP bound'). This bound applies to block codes transmitted over memoryless symmetric channels, and it significantly improves the tightness of the bounding techniques in [87] and [109], especially for codes of short to moderate block lengths (note, however, that the classical bound in [87] holds regardless of the channel symmetry). The key factor behind this improvement is the application of the channel symmetry to sidestep the intermediate stages of analyzing the *maximal* error probability of *fixed composition* codes as in [87] and [109]. Hence, the derivation in Chapter 5 directly considers the *average* error probability of *arbitrary* block codes. The ISP bound is applied to the BEC and to M-ary phase shift keying (PSK) modulated signals transmitted over the i.i.d. Rayleigh-fading and AWGN channels. For the latter channel, its tightness is also compared with the SP59 bound. The numerical

instability of existing algorithms for the numerical calculation of the SP59 for codes of moderate to large block lengths motivates the derivation of an alternative algorithm in Section 5.4 which facilitates the exact calculation of the this bound, irrespectively of the block length. Chapter 5 is a preprint of the submitted paper [119] (parts of these results were also published in the conference papers [82, 117]).

Chapter 2

Parity-Check Density versus Performance of Binary Linear Block Codes over Memoryless Symmetric Channels: New Bounds and Applications

This chapter is a reprint of

- 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, no. 2 pp. 550-579, February 2007.

Chapter Overview: The moderate complexity of low-density parity-check (LDPC) codes under iterative decoding is attributed to the sparseness of their parity-check matrices. It is therefore of interest to consider how sparse parity-check matrices of binary linear block codes can be as a function of their achievable rates and their gap to capacity. The remarkable performance of LDPC codes under practical and sub-optimal decoding algorithms makes it also interesting to investigate the inherent loss in performance which is attributed to the sub-optimality of iterative decoding, as well as the limitation imposed by the structure of the code. This paper addresses these two questions by introducing upper bounds on the achievable rates of binary linear block codes under maximum-likelihood (ML) decoding, and lower bounds on the asymptotic density of their parity-check matrices as a function of the achievable gap (in rate) to capacity; these bounds assume that the transmission takes place over a

memoryless binary-input output-symmetric channel. The new bounds improve some previously reported results, and are applied to ensembles of LDPC codes. The upper bounds on the achievable rates enable to assess the inherent gap in rate to capacity due to the structure of the ensemble, where this gap cannot be reduced even under ML decoding. The lower bounds on the asymptotic parity-check density are helpful in assessing the tradeoff between the asymptotic performance of LDPC codes and their decoding complexity (per iteration) under message-passing decoding.

2.1 Introduction

Error-correcting codes which employ iterative decoding algorithms are now considered state of the art in the field of low-complexity coding techniques. In [43], Khandekar and McEliece have suggested to study the encoding and decoding complexities of ensembles of iteratively decoded codes on graphs as a function of the gap between their achievable rates and capacity. They conjectured that if the achievable rate under iterative message-passing decoding is a fraction $1 - \varepsilon$ of the channel capacity, then for a wide class of channels, the encoding complexity scales like $\ln \frac{1}{\varepsilon}$ and the decoding complexity scales like $\frac{1}{\varepsilon} \ln \frac{1}{\varepsilon}$. The only exception is the binary erasure channel (BEC) where the decoding complexity behaves like $\ln \frac{1}{\varepsilon}$ (same as encoding complexity) due to the absolute reliability of the messages passed through the edges of the graph (hence, every edge can be used only once during the iterative decoding).

Low-density parity-check (LDPC) codes are efficiently decoded due to the sparseness of their parity-check matrices. In his thesis [30], Gallager proved that right-regular LDPC codes (i.e., LDPC codes with a constant degree (a_{R}) of the parity-check nodes) cannot achieve the channel capacity on a binary symmetric channel (BSC), even under maximum-likelihood (ML) decoding. This inherent gap to capacity is well approximated by an expression which decreases to zero exponentially fast in a_{R}. Richardson et al. [71] have extended this result, and proved that the same conclusion holds if a_{R} designates the *maximal right degree*. Sason and Urbanke later observed in [81] that the result still holds when considering the *average right degree*. Gallager's bound [30, Theorem 3.3] provides an upper bound on the rate of right-regular LDPC codes which achieve reliable communications over the BSC. Burshtein et al. have generalized Gallager's bound for general ensembles of LDPC codes transmitted over memoryless binary-input output-symmetric (MBIOS) channels [17]; to this end, they applied a two-level quantization to the log-likelihood ratio (LLR) of these channels which essentially turns them into a BSC.

Consider the number of ones in a parity-check matrix which represents a binary

linear block code, and normalize it per information bit (i.e., with respect to the dimension of the code). This quantity (which will be later defined as the *density* of the parity-check matrix) is equal to the normalized number of left to right (or right to left) messages per information bit which are passed in the corresponding bipartite graph during a single iteration of the message-passing decoder. In [81], Sason and Urbanke considered how sparse parity-check matrices of binary linear block codes can be, as a function of their gap to capacity (where this gap depends in general on the channel and on the decoding algorithm). An information-theoretic lower bound on the asymptotic density of parity-check matrices was derived in [81, Theorem 2.1] where this bound applies to every MBIOS channel and *every* sequence of binary linear block codes achieving a fraction $1-\varepsilon$ of the channel capacity with vanishing bit error probability. It holds for an arbitrary representation of these codes by full-rank parity-check matrices, and is of the form $\frac{K_1+K_2 \ln \frac{1}{\varepsilon}}{1-\varepsilon}$ where K_1 and K_2 are constants which only depend on the channel. Though the logarithmic behavior of this lower bound is in essence correct (due to a logarithmic behavior of the upper bound on the asymptotic parity-check density in [81, Theorem 2.2]), the lower bound in [81, Theorem 2.1] is *not* tight (with the exception of the BEC, as demonstrated in [81, Theorem 2.3], and possibly also the BSC). The derivation of the bounds in this paper was motivated by the desire to improve the results in [17, Theorems 1 and 2] and [81, Theorem 2.1] which are based on a two-level quantization of the LLR. The new bounding techniques introduced in this paper provide new upper bounds on the achievable rates of LDPC codes over MBIOS channels, and new lower bounds on their asymptotic parity-check density.

In [60], Measson et al. derived an upper bound on the thresholds under ML decoding of LDPC ensembles transmitted over the BEC. Their general approach relies on extrinsic information transfer (EXIT) charts, having a surprising and deep connection with the maximum a posteriori (MAP) threshold due to the area theorem for the BEC. Generalized extrinsic information transfer (GEXIT) charts were recently introduced by Measson et al. [58]; GEXIT charts form a generalization of the concept of EXIT charts, and they satisfy the area theorem for an arbitrary MBIOS channel (see [74, Section 3.4.10]). This conservation law enables one to get upper bounds on the thresholds of turbo-like ensembles under bit-MAP decoding. The bound was shown to be tight for the BEC [60], and is conjectured to be tight in general for MBIOS channels [58].

A new method for analyzing LDPC codes and low-density generator-matrix (LDGM) codes under bit-MAP decoding is introduced by Montanari in [62]. The method is based on a rigorous approach to spin glasses, and allows a construction of lower bounds

on the entropy of the transmitted message conditioned on the received one. The calculation of this bound is rather complicated, and its complexity grows exponentially with the maximal right and left degrees (see [62, Eqs. (6.2) and (6.3)]); this imposes a considerable difficulty in its calculation (especially, for continuous-output channels). Since the bounds in [60, 62] are derived for ensembles of codes, they are probabilistic in their nature; based on concentration arguments, they hold asymptotically in probability 1 as the block length goes to infinity. Based on heuristic statistical mechanics calculations, it was conjectured that the bounds in [62], which hold for general LDPC and LDGM ensembles over MBIOS channels, are tight.

We derive in this paper new bounds on the achievable rates and the asymptotic parity-check density of sequences of binary linear block codes. These bounds, which are efficiently calculated in software, apply to arbitrary sequences of codes whose transmission takes place over an MBIOS channel. It is emphasized that the information-theoretic bounds in [17, 81] and this paper are valid for *every* sequence of binary linear block codes, in contrast to high probability results. As examples for the latter category of probabilistic bounds which apply to ensembles, the reader is referred to the recent bounds of Montanari [62] under MAP decoding, the bound of Measson et al. for the BEC under MAP decoding [60], and the previously derived bound of Shokrollahi, relying on density evolution analysis for the BEC [93]. Shokrollahi proved in [93] that when the codes are communicated over a BEC, the growth rate of the average right degree (i.e., the average degree of the parity-check nodes in a bipartite Tanner graph) is at least logarithmic in terms of the gap to capacity. The statement in [93] is a high probability result which assumes a sub-optimal (iterative) decoding algorithm, whereas the statements in [17, 81] and this paper are valid even under ML decoding. As mentioned above, the bounds in [60, 62] refer to MAP decoding, but they form high probability results as the block length gets large.

The significance of the bounds in this paper is demonstrated in two respects. The new upper bounds on the achievable rates of binary linear block codes tighten previously reported bounds by Burshtein et al. [17]; therefore, they enable to obtain tighter upper bounds on the thresholds of sequences of binary linear block codes under ML decoding. These bounds are applied to LDPC codes, and the improvement in their tightness is exemplified numerically. Comparing the new upper bounds on the achievable rates with thresholds provided by a density-evolution analysis gives rigorous bounds on the inherent loss in performance due to the sub-optimality of message-passing decoding (as compared to soft-decision ML decoding), and also enables to assess the limitation imposed by the structure of the codes (or ensembles). The new lower bounds on the asymptotic parity-check density tighten the lower bound

in [81, Theorem 2.1]. Since the parity-check density can be interpreted as the complexity per iteration under message-passing decoding, then tightening the reported lower bound on the parity-check density [81] gives insight on the tradeoff between the asymptotic performance and decoding complexity of LDPC codes.

In this paper, preliminary material is presented in Section 2.2, and the theorems are introduced and proved in Sections 2.3 and 2.4. The derivation of the bounds in Section 2.3 was motivated by the desire to generalize the results in [17, Theorems 1 and 2] and [81, Theorem 2.1]. A two-level quantization of the LLR, in essence replacing the arbitrary MBIOS channel by a physically degraded BSC, is modified in Section 2.3 to a quantized channel which better reflects the statistics of the original channel (though the quantized channel is still physically degraded w.r.t. the original channel). The number of quantization levels of the LLR for the new channel is an arbitrary integer power of 2, and the calculation of these bounds is subject to an optimization of the quantization levels, as to obtain the tightest bounds within their form. In Section 2.4, the analysis relies on the conditional pdf of the LLR at the output of an MBIOS channel, and operates on an equivalent channel without quantizing the LLR. This second approach leads in Section 2.4 to bounds which are uniformly tighter than the bounds derived in Section 2.3 and are easier to calculate. The significance of the quantized and un-quantized bounds in Sections 2.3 and 2.4, respectively, stems from a comparison between these bounds which gives insight on the effect of the number of quantization levels of the LLR (even if they are optimally determined) on the achievable rates, as compared to the ideal case where no quantization is done. Numerical results are exemplified in Section 2.5. Finally, in Section 2.6, we summarize and present interesting issues which deserve further research. Four appendices provide further technical details referring to the proofs in Sections 2.3 and 2.4.

We note that the statements in this paper refer to the case where the parity-check matrices are full rank. Though it seems like a mild requirement for specific linear codes, this poses a problem when considering ensembles of LDPC codes. In the latter case, a parity-check matrix, referring to a randomly chosen bipartite graph with a given pair of degree distributions, may not be full rank.[1] Fortunately, as we later explain in this paper (see Section 2.5), the statements still hold for ensembles when we replace the code rate with the design rate.

[1]One can construct LDPC ensembles where the design rate is strictly less than the asymptotic rate as the block length goes to infinity; this can be done by simply repeating a non-vanishing fraction of the rows of a parity-check matrix, so that the design rate becomes strictly less than the rate, regardless of the block length.

2.2 Preliminaries

We introduce here some definitions and theorems from [17, 81] which serve as preliminary material for the rest of the paper. Definitions 2.1 and 2.2 are taken from [81, Section 2].

Definition 2.1 [Capacity-Approaching Codes] Let $\{\mathcal{C}_m\}$ be a sequence of codes, and denote the rate of the code $\mathcal{C}_m$ by R_m. Assume that for every m, the codewords of the code $\mathcal{C}_m$ are transmitted with equal probability over a channel whose capacity is C. This sequence is said to *achieve a fraction* $1-\varepsilon$ *of the channel capacity with vanishing bit error probability* if $\lim_{m\to\infty} R_m = (1-\varepsilon)C$, and there exists a decoding algorithm under which the average bit error probability of the code $\mathcal{C}_m$ tends to zero in the limit where $m \to \infty$.

Definition 2.2 [Parity-Check Density] Let $\mathcal{C}$ be a binary linear code of rate R and block length n, which is represented by a parity-check matrix H. We define the *density* of H, call it $\Delta = \Delta(H)$, as the normalized number of ones in H *per information bit*. The total number of ones in H is therefore equal to $nR\Delta$.

Definition 2.3 [Log-Likelihood Ratio (LLR)] Let us consider an MBIOS channel whose conditional pdf is $p_{Y|X}$ where X and Y designate the channel input and output, respectively. The log-likelihood ratio (LLR) at the output of the channel is

$$\mathrm{LLR}(y) \triangleq \ln\left(\frac{p_{Y|X}(y|0)}{p_{Y|X}(y|1)}\right).$$

Throughout the paper, we assume that all the codewords of a binary linear block code are equally likely to be transmitted. Also, the function h_2 designates the binary entropy function to base 2, i.e., $h_2(x) = -x\log_2(x) - (1-x)\log_2(1-x)$.

Theorem 2.1 [An Upper Bound on the Achievable Rates for Reliable Communication over MBIOS Channels] [17, Theorem 2]: Consider a sequence $\{\mathcal{C}_m\}$ of binary linear block codes of rate R_m, and assume that their block length tends to infinity as $m \to \infty$. Let H_m be a full-rank parity-check matrix of the code $\mathcal{C}_m$, and assume that $\Gamma_{k,m}$ designates the fraction of the parity-check equations involving k variables. Let

$$\Gamma_k \triangleq \lim_{m\to\infty} \Gamma_{k,m}, \quad R \triangleq \lim_{m\to\infty} R_m \tag{2.1}$$

where these limits are assumed to exist. Suppose that the transmission of these codes takes place over an MBIOS channel with capacity C bits per channel use, and let

$$w \triangleq \frac{1}{2}\int_{-\infty}^{\infty} \min\big(f(y), f(-y)\big)\,\mathrm{d}y \tag{2.2}$$

where $f(y) \triangleq p_{Y|X}(y|0)$ designates the conditional pdf of the output of the MBIOS channel when zero is transmitted. Then, a necessary condition for vanishing block error probability as $m \to \infty$ is

$$R \leq 1 - \frac{1-C}{\sum_k \left\{ \Gamma_k \, h_2\left(\frac{1-(1-2w)^k}{2}\right)\right\}}.$$

Theorem 2.2 [Lower Bounds on the Asymptotic Parity-Check Density with Two-Level Quantization] [81, Theorem 2.1]: Let $\{\mathcal{C}_m\}$ be a sequence of binary linear block codes achieving a fraction $1-\varepsilon$ of the capacity of an MBIOS channel with vanishing bit error probability. Denote Δ_m as the density of a full-rank parity-check matrix of the code $\mathcal{C}_m$. Then, the asymptotic density satisfies

$$\liminf_{m\to\infty} \Delta_m > \frac{K_1 + K_2 \ln \frac{1}{\varepsilon}}{1-\varepsilon} \tag{2.3}$$

where

$$K_1 = \frac{(1-C)\,\ln\left(\frac{1}{2\ln 2}\,\frac{1-C}{C}\right)}{2C\,\ln\left(\frac{1}{1-2w}\right)}, \quad K_2 = \frac{1-C}{2C\,\ln\left(\frac{1}{1-2w}\right)} \tag{2.4}$$

and w is defined in (2.2). For a BEC with erasure probability p, the coefficients K_1 and K_2 in (2.4) are improved to

$$K_1 = \frac{p\,\ln\left(\frac{p}{1-p}\right)}{(1-p)\,\ln\left(\frac{1}{1-p}\right)}, \quad K_2 = \frac{p}{(1-p)\,\ln\left(\frac{1}{1-p}\right)}. \tag{2.5}$$

The bounds in this paper are applied to low-density parity-check (LDPC) codes. In general, LDPC codes are linear block codes which are represented by a sparse parity-check matrix H. This matrix can be represented in an equivalent form by a bipartite graph $\mathcal{G}$ whose variable nodes (appearing on the left of $\mathcal{G}$) represent the code bits, and whose parity-check nodes (appearing on the right of $\mathcal{G}$) represent the linear constraints defined by H. In such a bipartite graph, an edge connects a variable node with a parity-check node if and only if the corresponding code bit is involved in the parity-check equation; the degree of a node is defined as the number of edges which are adjacent to it.

Following standard notation, let λ_i and ρ_i denote the fraction of *edges* attached to variable and parity-check nodes of degree i, respectively. In a similar manner, let Λ_i and Γ_i denote the fraction of variable and parity-check nodes of degree i, respectively. The LDPC ensemble is characterized by a triple (n, λ, ρ) where n designates the block length of the codes, and the polynomials

$$\lambda(x) \triangleq \sum_{i=1}^{\infty} \lambda_i x^{i-1}, \qquad \rho(x) \triangleq \sum_{i=1}^{\infty} \rho_i x^{i-1}$$

represent, respectively, the left and right degree distributions (d.d.) from the *edge* perspective. Equivalently, this ensemble can be also characterized by the triple LDPC(n, Λ, Γ) where the polynomials

$$\Lambda(x) \triangleq \sum_{i=1}^{\infty} \Lambda_i x^i, \qquad \Gamma(x) \triangleq \sum_{i=1}^{\infty} \Gamma_i x^i$$

represent, respectively, the left and right d.d. from the *node* perspective. We denote by LDPC(n, λ, ρ) (or LDPC(n, Λ, Γ)) the ensemble of codes whose bipartite graphs are constructed according to the corresponding pairs of degree distributions. One can switch between degree distributions w.r.t. to the nodes and edges of a bipartite graph, using the following equations [74]:

$$\Lambda(x) = \frac{\int_0^x \lambda(u)du}{\int_0^1 \lambda(u)\mathrm{d}u}, \qquad \Gamma(x) = \frac{\int_0^x \rho(u)du}{\int_0^1 \rho(u)\mathrm{d}u} \tag{2.6}$$

$$\lambda(x) = \frac{\Lambda'(x)}{\Lambda'(1)}, \qquad \rho(x) = \frac{\Gamma'(x)}{\Gamma'(1)}. \tag{2.7}$$

An important characteristic of an ensemble of LPDC codes is its *design rate*. For an LDPC ensemble whose codes are represented by parity-check matrices of dimension $c \times n$, the design rate is defined to be $R_{\mathrm{d}} \triangleq 1 - \frac{c}{n}$. This serves as a lower bound on the actual rate of any code from this ensemble, and is equal to the actual rate if the parity-check matrix of a code is *full rank* (i.e., the linear constraints which define this code are linearly independent). For an ensemble of LDPC codes, the design rate is given in terms of the degree distributions (either w.r.t. the edges or nodes of a graph), and it can be expressed in two equivalent forms:

$$R_{\mathrm{d}} = 1 - \frac{\int_0^1 \rho(x)\mathrm{d}x}{\int_0^1 \lambda(x)\mathrm{d}x} = 1 - \frac{\Lambda'(1)}{\Gamma'(1)}. \tag{2.8}$$

A sufficient condition for the asymptotic convergence of the rate of codes from an LDPC ensemble to its design rate was recently stated in [60, Lemma 7].

Lemma 2.1 [A sufficiency condition for the equality between the design rate and asymptotic rate for ensembles of LDPC codes] [60, Lemma 7]: Let $\mathcal{C}$ be a code which is chosen uniformly at random from the ensemble LDPC(n, Λ, Γ), let R be the rate of $\mathcal{C}$, and let R_{d} be the design rate of this ensemble. Consider the

function

$$\begin{aligned}\Psi_{(\Lambda,\Gamma)}(u) &\triangleq -\Lambda'(1)\log_2\left[\frac{1+uv}{(1+u)(1+v)}\right] \\ &\quad + \sum_{i=1}^{\infty}\Lambda_i \log_2\left[\frac{1+u^i}{2(1+u)^i}\right] + \frac{\Lambda'(1)}{\Gamma'(1)}\sum_{i=1}^{\infty}\Gamma_i\log_2\left[1+\left(\frac{1-v}{1+v}\right)^i\right]\end{aligned}$$

where

$$v \triangleq \left(\sum_{i=1}^{\infty}\frac{\lambda_i}{1+u^i}\right)^{-1}\left(\sum_{i=1}^{\infty}\frac{\lambda_i u^{i-1}}{1+u^i}\right).$$

Assume that the function $\Psi_{(\Lambda,\Gamma)}$ achieves its global maximum in the range $u \in [0,\infty)$ at $u = 1$. Then, there exists a constant $B > 0$ such that, for any $\xi > 0$ and $n > n_0(\xi,\Lambda,\Gamma)$,

$$\Pr\{|R - R_{\mathrm{d}}| > \xi\} \leq e^{-Bn\xi}.$$

Moreover, there exists $C > 0$ such that, for $n > n_0(\xi,\Lambda,\Gamma)$

$$\mathbb{E}\{|R - R_{\mathrm{d}}|\} \leq \frac{C\log n}{n}.$$

In Section 2.5, we rely on this lemma in order to verify that the asymptotic rates of codes randomly chosen (with uniform distribution) from various ensembles of LDPC codes tend in probability 1 to the design rates of these ensembles.

2.3 Approach I: Bounds Based on Quantization of the LLR

In this section, we introduce bounds on the achievable rates and the asymptotic parity-check density of sequences of binary linear block codes. The bounds generalize previously reported results in [17] and [81] which were based on a symmetric two-level quantization of the LLR. This is achieved by extending the concept of quantization to an arbitrary integer power of 2; to this end, the analysis relies on the Galois field $\mathrm{GF}(2^d)$ where $d \in \mathbb{N}$. In Section 2.3.1, we demonstrate the results and their proofs for a four-level quantization. In Section 2.3.2, we extend the results to a symmetric quantization with a number of levels which is an arbitrary integer power of 2. This order of presentation was chosen since many concepts which are helpful for the generalization in Section 2.3.2 are written in a simplified notation for the four-level quantization, along with all the relevant lemmas for the general case which are already introduced in the derivation of the bound with four-level quantization. This also shortens considerably the proof for the general quantization in Section 2.3.2.

2.3.1 Bounds for Four-Levels of Quantization

As a preparatory step towards developing bounds on the parity-check density and the rate of binary linear block codes, we present a lower bound on the conditional entropy of a transmitted codeword given the received sequence at the output of an arbitrary MBIOS channel.

Proposition 2.1 Let $\mathcal{C}$ be a binary linear block code of length n and rate R, and assume that its transmission takes place over an MBIOS channel whose conditional pdf is given by $p_{Y|X}$. Let $\mathbf{X} = (X_1, \ldots, X_n)$ and $\mathbf{Y} = (Y_1, \ldots, Y_n)$ designate the transmitted codeword and received sequence, respectively. For an arbitrary positive $l \in \mathbb{R}^+$, let us define the probabilities p_0, p_1, p_2, p_3 as follows:

$$\begin{aligned}
p_0 &\triangleq \Pr\{\mathrm{LLR}(Y) > l \,|\, X = 0\} \\
p_1 &\triangleq \Pr\{\mathrm{LLR}(Y) \in (0, l] \,|\, X = 0\} + \frac{1}{2}\Pr\{\mathrm{LLR}(Y) = 0 \,|\, X = 0\} \\
p_2 &\triangleq \Pr\{\mathrm{LLR}(Y) \in [-l, 0) \,|\, X = 0\} + \frac{1}{2}\Pr\{\mathrm{LLR}(Y) = 0 \,|\, X = 0\} \\
p_3 &\triangleq \Pr\{\mathrm{LLR}(Y) < -l \,|\, X = 0\}.
\end{aligned} \tag{2.9}$$

For an arbitrary full-rank parity-check matrix of the code $\mathcal{C}$, let Γ_k designate the fraction of parity-check equations involving k variables. Then, the conditional entropy of the transmitted codeword given the received sequence satisfies

$$\begin{aligned}
\frac{H(\mathbf{X}|\mathbf{Y})}{n} \geq 1 - C - (1 - R) \cdot \\
\cdot \sum_k \left\{ \Gamma_k \sum_{t=0}^{k} \binom{k}{t} (p_1 + p_2)^t (p_0 + p_3)^{k-t} \right. \\
\left. \cdot h_2 \left(\frac{1 - \left(1 - \frac{2p_2}{p_1+p_2}\right)^t \left(1 - \frac{2p_3}{p_0+p_3}\right)^{k-t}}{2} \right) \right\}.
\end{aligned} \tag{2.10}$$

Remark 2.1 Note that the input vector $\mathbf{X}$ is chosen uniformly from the codewords of a binary linear block code. Each input bit X_i therefore either gets the values 0 or 1 with probability $\frac{1}{2}$ or is set to zero (due to the linearity of the code). In the following proof, we assume that all the code symbols get the values 0 or 1 with equal probability. By slightly modifying the proof, it is simple to show that the bound also holds for the other case where some of the code bits are set to zero. Without mentioning explicitly, the same assumption will be taken in the proofs of Propositions 2.2 and 2.3.

Proof: Considering an MBIOS channel whose conditional pdf is given by $p_{Y|X}$, we introduce a new physically degraded channel. It is a binary-input, quaternary-output

symmetric channel (see Figure 2.1). To this end, let $l \in \mathbb{R}^+$ be an arbitrary positive number, and let α be a primitive element of the Galois field GF(2^2) (so $\alpha^2 = 1 + \alpha$). The set of the elements of this field is $\{0, 1, \alpha, 1 + \alpha\}$. Let X_i and Y_i designate the random variables referring to the input and output of the original channel at time i (where $i = 1, 2, \ldots, n$). We define the degraded channel as a channel with four quantization levels of the LLR. The output of the degraded channel at time i, Z_i, is calculated from the output Y_i of the original channel as follows:

- If $\mathrm{LLR}(Y_i) > l$, then $Z_i = 0$.
- If $0 < \mathrm{LLR}(Y_i) \leq l$, then $Z_i = \alpha$.
- If $-l \leq \mathrm{LLR}(Y_i) < 0$, then $Z_i = 1 + \alpha$.
- If $\mathrm{LLR}(Y_i) < -l$, then $Z_i = 1$.
- If $\mathrm{LLR}(Y_i) = 0$, then Z_i is chosen as α or $1 + \alpha$ with equal probability ($\frac{1}{2}$).

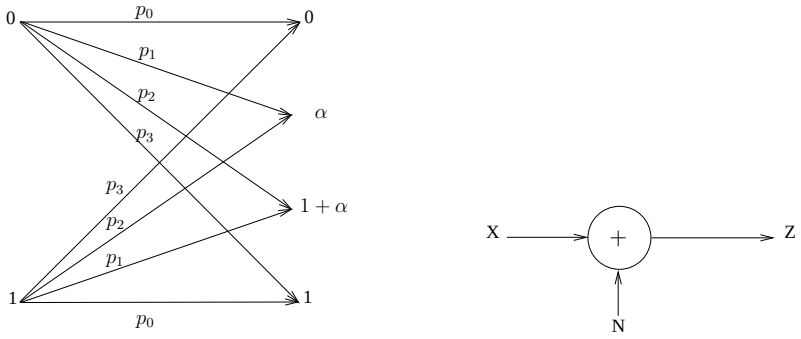

Figure 2.1: The channel model in the left plot is a physically degraded channel used for the derivation of the bound with four levels of quantization. The element α denotes a primitive element in GF(2^2). This channel model is equivalent to a channel with an additive noise in GF(2^2) (see right plot).

From the definition of the degraded channel in Figure 2.1, this channel has an additive noise in GF(2^2) and is also binary-input output-symmetric. It follows that the transition probabilities of the degraded channel are

$$\begin{aligned}
p_0 &= \Pr(Z = 0 \,|\, X = 0) = \Pr(Z = 1 \,|\, X = 1) \\
p_1 &= \Pr(Z = \alpha \,|\, X = 0) = \Pr(Z = 1 + \alpha \,|\, X = 1) \\
p_2 &= \Pr(Z = 1 + \alpha \,|\, X = 0) = \Pr(Z = \alpha \,|\, X = 1) \\
p_3 &= \Pr(Z = 1 \,|\, X = 0) = \Pr(Z = 0 \,|\, X = 1)
\end{aligned}$$

where p_j is introduced in (2.9) for $0 \leq j \leq 3$, and the symmetry in these transition probabilities holds since the original channel is MBIOS.

Since $\mathcal{C}$ is a binary linear block code of length n and rate R, and the codewords are transmitted with equal probability then

$$H(\mathbf{X}) = nR. \tag{2.11}$$

Also, since the channel is memoryless, then

$$H(\mathbf{Y}|\mathbf{X}) = nH(Y|X). \tag{2.12}$$

We designate the output sequence of the degraded channel by $\mathbf{Z} = (Z_1, \ldots, Z_n)$. Since the mapping from Y_i to the degraded output Z_i $(i = 1, 2, \cdots, n)$ is memoryless, then $H(\mathbf{Z}|\mathbf{Y}) = nH(Z|Y)$ and

$$\begin{aligned} H(\mathbf{Y}) &= H(\mathbf{Z}) - H(\mathbf{Z}|\mathbf{Y}) + H(\mathbf{Y}|\mathbf{Z}) \\ &= H(\mathbf{Z}) - nH(Z|Y) + H(\mathbf{Y}|\mathbf{Z}) \end{aligned} \tag{2.13}$$

$$\begin{aligned} H(\mathbf{Y}|\mathbf{Z}) &\leq \sum_{i=1}^{n} H(Y_i|Z_i) \\ &= nH(Y|Z) \\ &= n\left[H(Y) - H(Z) + H(Z|Y)\right]. \end{aligned} \tag{2.14}$$

Applying the above towards a lower bound on the conditional entropy $H(\mathbf{X}|\mathbf{Y})$, we get

$$\begin{aligned} H(\mathbf{X}|\mathbf{Y}) &= H(\mathbf{X}) + H(\mathbf{Y}|\mathbf{X}) - H(\mathbf{Y}) \\ &= nR + nH(Y|X) - H(\mathbf{Y}) \\ &= nR + nH(Y|X) - H(\mathbf{Z}) - H(\mathbf{Y}|\mathbf{Z}) + nH(Z|Y) \\ &\geq nR + nH(Y|X) - H(\mathbf{Z}) - n\left[H(Y) - H(Z) + H(Z|Y)\right] + nH(Z|Y) \\ &= nR - H(\mathbf{Z}) + nH(Z) - n\left[H(Y) - H(Y|X)\right] \\ &= nR - H(\mathbf{Z}) + nH(Z) - nI(X;Y) \\ &\geq nR - H(\mathbf{Z}) + nH(Z) - nC \end{aligned} \tag{2.15}$$

where the second equality relies on (2.11) and (2.12), the third equality relies on (2.13), the first inequality relies on (2.14), and $I(X;Y) \leq C$ is used for the last transition (where C designates the capacity of the original channel).

In order to obtain a lower bound on $H(\mathbf{X}|\mathbf{Y})$ from (2.15), we calculate the entropy of the random variable Z, and derive an upper bound on the entropy of the random

vector $\mathbf{Z}$. This finally provides the lower bound in (2.10). Observing that the degraded channel is additive over GF(2^2), we denote the additive noise by

$$N_i = \Theta_i + \Omega_i \alpha, \quad i \in \{1, \ldots, n\}$$

where $\mathbf{\Theta} = (\Theta_1, \ldots, \Theta_n)$ and $\mathbf{\Omega} = (\Omega_1, \ldots, \Omega_n)$ are random vectors over GF(2). Note that $\mathbf{\Theta}$ and $\mathbf{\Omega}$ are statistically independent of the transmitted codeword $\mathbf{X}$. Since the code is binary, it follows that

$$Z_i = \Phi_i + \Omega_i \alpha \tag{2.16}$$

where $\Phi_i \triangleq \Theta_i + X_i$. This gives

$$\begin{aligned} H(Z) &= H(\Phi, \Omega) \\ &= H(\Omega) + H(\Phi|\Omega) \\ &= H(\Omega) + 1 \end{aligned} \tag{2.17}$$

where the last equality follows since $\mathcal{C}$ is a binary linear block code which implies that the input X is equally likely to be zero or one; since Ω is independent of X, then Φ is equally likely to be zero or one given the value of Ω.

We now derive an upper bound on the entropy $H(\mathbf{Z})$. Based on (2.16), it is easy to verify the following chain of equalities:

$$\begin{aligned} H(\mathbf{Z}) &= H(Z_1, \ldots, Z_n) \\ &= H(\Phi_1, \ldots, \Phi_n, \Omega_1, \ldots, \Omega_n) \\ &= H(\Omega_1, \ldots, \Omega_n) + H(\Phi_1, \ldots, \Phi_n \,|\, \Omega_1, \ldots, \Omega_n) \\ &= n\, H(\Omega) + H(\Phi_1, \ldots, \Phi_n \,|\, \Omega_1, \ldots, \Omega_n) \end{aligned} \tag{2.18}$$

where the last equality follows since the degraded channel in Figure 2.1 is memoryless. Let us define the syndrome at the output of the degraded channel as

$$\mathbf{S} \triangleq (\Phi_1, \ldots, \Phi_n)\, H^T$$

where H is a full-rank parity-check matrix of the binary linear block code $\mathcal{C}$. We note that the calculation of the syndrome only takes into account the Φ-components of the vector $\mathbf{Z}$ in (2.16). Also note that since $\mathbf{X}\, H^T = 0$ for every codeword $\mathbf{X}$, then $\mathbf{S} = (\Theta_1, \ldots, \Theta_n)\, H^T$ which is independent of the transmitted codeword. Let us define M as the index of the vector $(\Phi_1, \ldots, \Phi_n)$ in the coset referring to the syndrome $\mathbf{S}$. Since each coset has exactly 2^{nR} elements which are equally likely, then $H(M) = nR$, and

$$\begin{aligned} H(\Phi_1, \ldots, \Phi_n \,|\, \Omega_1, \ldots, \Omega_n) &= H(\mathbf{S}, M \,|\, \Omega_1, \ldots, \Omega_n) \\ &\leq H(M) + H(\mathbf{S} \,|\, \Omega_1, \ldots, \Omega_n) \\ &= nR + H(\mathbf{S} \,|\, \Omega_1, \ldots, \Omega_n). \end{aligned} \tag{2.19}$$

Considering a parity-check equation involving k variables, let $\{i_1, \ldots, i_k\}$ be the set of indices of the variables involved in this parity-check equation. The relevant component of the syndrome $\mathbf{S}$ which refers to this parity-check equation is equal to zero or one if and only if the Hamming weight sub-vector $(\Theta_{i_1}, \ldots, \Theta_{i_k})$ is even or odd, respectively. It is clear from Figure 2.1 that for an index i for which $\Omega_i = 1$, Θ_i is equal to one in probability $\frac{p_2}{p_1+p_2}$. Similarly, for an index i for which $\Omega_i = 0$, then Θ_i is equal to one in probability $\frac{p_3}{p_0+p_3}$.

Given that the Hamming weight of the vector $(\Omega_{i_1}, \ldots, \Omega_{i_k})$ is t, then the probability of an even Hamming weight of the random vector $(\Theta_{i_1}, \ldots, \Theta_{i_k})$ is equal to

$$q_1(t,k)\, q_2(t,k) + \big(1 - q_1(t,k)\big)\, \big(1 - q_2(t,k)\big)$$

where $q_1(t,k)$ designates the probability that among the t indices i for which $\Omega_i = 1$, the random variable Θ_i is equal to 1 an even number of times, and $q_2(t,k)$ designates the probability that the same happens for the $k-t$ indices i for which $\Phi_i = 0$. Based on the discussion above, it follows that

$$q_1(t,k) = \sum_{i \,\text{even}} \left\{ \binom{t}{i} \left(\frac{p_1}{p_1+p_2}\right)^{t-i} \left(\frac{p_2}{p_1+p_2}\right)^{i} \right\} = \frac{1 + \left(1 - \frac{2p_2}{p_1+p_2}\right)^t}{2}$$

$$q_2(t,k) = \sum_{i \,\text{even}} \left\{ \binom{k-t}{i} \left(\frac{p_0}{p_0+p_3}\right)^{k-t-i} \left(\frac{p_3}{p_0+p_3}\right)^{i} \right\} = \frac{1 + \left(1 - \frac{2p_3}{p_0+p_3}\right)^{k-t}}{2}.$$

Hence, the probability that the vector $(\Theta_{i_1}, \Theta_{i_2}, \ldots, \Theta_{i_k})$ is of even Hamming weight is

$$q_1(t,k)\, q_2(t,k) + \big(1 - q_1(t,k)\big)\, \big(1 - q_2(t,k)\big) = \frac{1 + \left(1 - \frac{2p_2}{p_1+p_2}\right)^t \left(1 - \frac{2p_3}{p_0+p_3}\right)^{k-t}}{2}.$$

We conclude that given a vector $\omega \in \{0,1\}^k$ of Hamming weight t

$$H\big(S_i \,|\, (\Omega_{i_1}, \ldots, \Omega_{i_k}) = \omega\big) = h_2\left(\frac{1 + \left(1 - \frac{2p_2}{p_1+p_2}\right)^t \left(1 - \frac{2p_3}{p_0+p_3}\right)^{k-t}}{2} \right).$$

This yields that if the calculation of a component S_i $(i = 1, \ldots, n(1-R))$ in the syndrome $\mathbf{S}$ relies on a parity-check equation involving k variables, then

$$\begin{aligned} &H(S_i \,|\, \Omega_1, \ldots, \Omega_n) \\ &= H(S_i \,|\, \Omega_{i_1}, \ldots, \Omega_{i_k}) \\ &= \sum_{\omega \in \{0,1\}^k} \Pr\big((\Omega_{i_1}, \ldots, \Omega_{i_k}) = \omega\big) \cdot H\left(S_i \,|\, (\Omega_{i_1}, \ldots, \Omega_{i_k}) = \omega\right) \end{aligned}$$

$$= \sum_{t=0}^{k} \binom{k}{t} (p_1+p_2)^t (p_0+p_3)^{k-t} h_2 \left(\frac{1 + \left(1 - \frac{2p_2}{p_1+p_2}\right)^t \left(1 - \frac{2p_3}{p_0+p_3}\right)^{k-t}}{2} \right)$$

$$= \sum_{t=0}^{k} \binom{k}{t} (p_1+p_2)^t (p_0+p_3)^{k-t} h_2 \left(\frac{1 - \left(1 - \frac{2p_2}{p_1+p_2}\right)^t \left(1 - \frac{2p_3}{p_0+p_3}\right)^{k-t}}{2} \right)$$

where the third equality turns to averaging over the Hamming weight of $(\Omega_{i_1}, \ldots, \Omega_{i_k})$ (note that each component is Bernoulli distributed with $\Pr(\Omega_i = 0) = p_0 + p_3$), and the last equality follows from the symmetry of the binary entropy function (where $h_2(x) = h_2(1-x)$ for $x \in [0,1]$). Let Γ_k designate the fraction of parity-check equations in the full-rank parity-check matrix which involve k variables, so their total number is $n(1-R)\Gamma_k$ and

$$\begin{aligned} & H(\mathbf{S} \,|\, \Phi_1, \Phi_2, \ldots, \Phi_n) \\ & \leq \sum_{i=1}^{n(1-R)} H(S_i \,|\, \Phi_1, \Phi_2, \ldots, \Phi_n) \\ & = n(1-R) \sum_k \left\{ \Gamma_k \sum_{t=0}^{k} \binom{k}{t} (p_1+p_2)^t (p_0+p_3)^{k-t} \right. \\ & \qquad \left. \cdot h_2 \left(\frac{1 - \left(1 - \frac{2p_2}{p_1+p_2}\right)^t \left(1 - \frac{2p_3}{p_0+p_3}\right)^{k-t}}{2} \right) \right\}. \end{aligned} \tag{2.20}$$

By combining (2.18)–(2.20), an upper bound on the entropy of the random vector $\mathbf{Z}$ follows:

$$\begin{aligned} H(\mathbf{Z}) \;\leq\; & nR + nH(\Omega) \\ & + n(1-R) \sum_k \left\{ \Gamma_k \sum_{t=0}^{k} \binom{k}{t} (p_1+p_2)^t (p_0+p_3)^{k-t} \right. \\ & \qquad \left. \cdot h_2 \left(\frac{1 - \left(1 - \frac{2p_2}{p_1+p_2}\right)^t \left(1 - \frac{2p_3}{p_0+p_3}\right)^{k-t}}{2} \right) \right\}. \end{aligned} \tag{2.21}$$

The substitution of (2.17) and (2.21) in (2.15) finally provides the lower bound on the conditional entropy $H(\mathbf{X} \,|\, \mathbf{Y})$ in (2.10). ■

The following theorem tightens the lower bound on the parity-check density of an arbitrary sequence of binary linear block codes given in [81, Theorem 2.1]. It is based on a four-level quantization of the LLR at the output of an MBIOS channel (as opposed to the two-level quantization of the LLR used in [81]).

Theorem 2.3 [**"Four-Level Quantization" Lower Bound on the Asymptotic Parity-Check Density of Binary Linear Block Codes**] Let $\{C_m\}$ be a sequence of binary linear block codes achieving a fraction $1-\varepsilon$ of the capacity of an MBIOS channel with vanishing bit error probability. Let H_m be an arbitrary *full-rank* parity-check matrix of the code $\mathcal{C}_m$, and denote its density by Δ_m. Then, the asymptotic density satisfies

$$\liminf_{m\to\infty} \Delta_m > \frac{K_1 + K_2 \ln \frac{1}{\varepsilon}}{1-\varepsilon} \tag{2.22}$$

where

$$K_1 = K_2 \, \ln\left(\frac{1}{2\ln 2}\,\frac{1-C}{C}\right), \quad K_2 = -\frac{1-C}{C \, \ln\left(\frac{(p_1-p_2)^2}{p_1+p_2} + \frac{(p_0-p_3)^2}{p_0+p_3}\right)} \tag{2.23}$$

and p_0, p_1, p_2, p_3 are defined in (2.9) in terms of $l \in \mathbb{R}^+$. The optimal value of l is given implicitly as a solution to the equation

$$\frac{p_2^2 + e^{-l} p_1^2}{(p_1+p_2)^2} = \frac{p_3^2 + e^{-l} p_0^2}{(p_0+p_3)^2} \tag{2.24}$$

where such a solution always exists.[2]

Proof: *Derivation of the lower bound in* (2.22) *and* (2.23):

Lemma 2.2 Let $\mathcal{C}$ be a binary linear block code of length n and rate R. Let P_{b} designate the average bit error probability of the code $\mathcal{C}$ which is associated with an arbitrary decoding algorithm and channel, and let $\mathbf{X}$ and $\mathbf{Y}$ designate the transmitted codeword and received sequence, respectively. Then

$$\frac{H(\mathbf{X}\,|\,\mathbf{Y})}{n} \leq R\, h_2(P_{\mathrm{b}}). \tag{2.25}$$

Proof: The lemma is proved in Appendix 2.A.1. ■

Lemma 2.3 $h_2(x) \leq 1 - \frac{2}{\ln 2}\,(\frac{1}{2} - x)^2$ for $0 \leq x \leq 1$.

Proof: The lemma is proved in [81, Lemma 3.1]; this inequality actually forms a particular case of Eq. (2.B.2) whose derivation is based on truncating the power series expansion of the binary entropy function around $\frac{1}{2}$. ■

Referring to an arbitrary sequence of binary linear block codes $\{\mathcal{C}_m\}$ which achieves a fraction $1-\varepsilon$ of capacity with vanishing bit error probability, then according to

[2] It was observed numerically that the solution l of the optimization equation (2.24) is unique when considering the binary-input AWGN channel. We conjecture that the uniqueness of such a solution is a property which holds for MBIOS channels under some mild conditions.

Definition 2.1, there exists a decoding algorithm (e.g., ML decoding) so that the average bit error probability of the code $\mathcal{C}_m$ tends to zero as m goes to infinity, and $\lim_{m\to\infty} R_m = (1-\varepsilon)C$. From Lemma 2.2, we obtain that $\lim_{m\to\infty} \frac{H(\mathbf{X_m}\,|\,\mathbf{Y_m})}{n_m} = 0$ where $\mathbf{X_m}$ and $\mathbf{Y_m}$ designate the transmitted codeword in the code $\mathcal{C}_m$ and the received sequence, respectively , and n_m designates the block length of the code $\mathcal{C}_m$. From Proposition 2.1, we obtain

$$\begin{aligned}\frac{H(\mathbf{X_m}|\mathbf{Y_m})}{n_m} \;\geq\; & 1 - C - (1-R_m) \\ & \cdot \sum_k \left\{ \Gamma_{k,m} \sum_{t=0}^{k} \binom{k}{t} (p_1+p_2)^t (p_0+p_3)^{k-t} \right. \\ & \left. \cdot h_2\left(\frac{1 - \left(1 - \frac{2p_2}{p_1+p_2}\right)^t \left(1 - \frac{2p_3}{p_0+p_3}\right)^{k-t}}{2} \right) \right\}\end{aligned}$$

where $\Gamma_{k,m}$ designates the fraction of parity-check equations in a parity-check matrix H_m which involve k variables. The upper bound on the binary entropy function h_2 in Lemma 2.3 gives

$$\begin{aligned}\frac{H(\mathbf{X_m}|\mathbf{Y_m})}{n_m} \;\geq\; & 1 - C - (1-R_m) \\ & \cdot \sum_k \left\{ \Gamma_{k,m} \sum_{t=0}^{k} \binom{k}{t} (p_1+p_2)^t (p_0+p_3)^{k-t} \right. \\ & \left. \left[1 - \frac{1}{2\ln 2} \left(\frac{p_1-p_2}{p_1+p_2} \right)^{2t} \left(\frac{p_0-p_3}{p_0+p_3} \right)^{2(k-t)} \right] \right\}\end{aligned} \tag{2.26}$$

Since $p_0+p_1+p_2+p_3 = 1$ (i.e., the transition probabilities of the degraded channel in Figure 2.1 sum to 1), then

$$\begin{aligned}& \sum_k \left\{ \Gamma_{k,m} \sum_{t=0}^{k} \binom{k}{t} (p_1+p_2)^t (p_0+p_3)^{k-t} \left[1 - \frac{1}{2\ln 2} \left(\frac{p_1-p_2}{p_1+p_2} \right)^{2t} \left(\frac{p_0-p_3}{p_0+p_3} \right)^{2(k-t)} \right] \right\} \\ & = \sum_k \left\{ \Gamma_{k,m} \left[1 - \frac{1}{2\ln 2} \sum_{t=0}^{k} \binom{k}{t} \left(\frac{(p_1-p_2)^2}{p_1+p_2} \right)^t \left(\frac{(p_0-p_3)^2}{p_0+p_3} \right)^{k-t} \right] \right\} \\ & = 1 - \frac{1}{2\ln 2} \sum_k \left\{ \Gamma_{k,m} \sum_{t=0}^{k} \binom{k}{t} \left(\frac{(p_1-p_2)^2}{p_1+p_2} \right)^t \left(\frac{(p_0-p_3)^2}{p_0+p_3} \right)^{k-t} \right\} \\ & = 1 - \frac{1}{2\ln 2} \sum_k \left\{ \Gamma_{k,m} \left(\frac{(p_1-p_2)^2}{p_1+p_2} + \frac{(p_0-p_3)^2}{p_0+p_3} \right)^k \right\} \\ & \leq 1 - \frac{1}{2\ln 2} \left(\frac{(p_1-p_2)^2}{p_1+p_2} + \frac{(p_0-p_3)^2}{p_0+p_3} \right)^{a_{\mathrm{R}}(m)}\end{aligned} \tag{2.27}$$

where $a_{\mathrm{R}}(m) \triangleq \sum_k k\Gamma_{k,m}$ designates the average right degree of the bipartite graph which refers to the parity-check matrix H_m, and the last transition follows from Jensen's inequality. Substituting (2.27) into the RHS of (2.26) and letting m tend to infinity give the inequality

$$1 - C - \big(1 - (1-\varepsilon)C\big) \left(1 - \frac{1}{2\ln 2} \left(\frac{(p_1 - p_2)^2}{p_1 + p_2} + \frac{(p_0 - p_3)^2}{p_0 + p_3}\right)^{a_{\mathrm{R}}(\infty)}\right) \leq 0 \quad (2.28)$$

where $a_{\mathrm{R}}(\infty) \triangleq \liminf_{m\to\infty} a_{\mathrm{R}}(m)$. Note that the base of the exponent in the LHS of this inequality does not exceed unity, i.e.,

$$\begin{aligned} &\frac{(p_1 - p_2)^2}{p_1 + p_2} + \frac{(p_0 - p_3)^2}{p_0 + p_3} \\ &\leq \frac{(p_1 + p_2)^2}{p_1 + p_2} + \frac{(p_0 + p_3)^2}{p_0 + p_3} \\ &= p_0 + p_1 + p_2 + p_3 = 1. \end{aligned}$$

Therefore, the inequality in (2.28) yields the following lower bound on the asymptotic average right degree:

$$a_{\mathrm{R}}(\infty) \geq K_1' + K_2' \ln\left(\frac{1}{\varepsilon}\right) \quad (2.29)$$

where

$$K_1' = -\frac{\ln\left(\frac{1}{2\ln 2}\frac{1-C}{C}\right)}{\ln\left(\frac{(p_1-p_2)^2}{p_1+p_2} + \frac{(p_0-p_3)^2}{p_0+p_3}\right)}, \quad K_2' = -\frac{1}{\ln\left(\frac{(p_1-p_2)^2}{p_1+p_2} + \frac{(p_0-p_3)^2}{p_0+p_3}\right)}. \quad (2.30)$$

According to Definition 2.2, the density (Δ) of a parity-check matrix is equal to the number of edges in the corresponding bipartite graph normalized per information bit, while the average right degree (a_{R}) is equal to the same number of edges normalized per parity-check node. Since the parity-check matrix H is full rank, then the above scalings of the number of edges in a bipartite graph imply

$$\Delta = \frac{1-R}{R}\, a_{\mathrm{R}} \quad (2.31)$$

where R is the rate of a binary linear block code. By our assumption, the asymptotic rate of the sequence of code $\{\mathcal{C}_m\}$ is equal to a fraction $1-\varepsilon$ of the capacity. Therefore, by combining (2.29) and (2.31) with $R = (1-\varepsilon)C$, we obtain a lower bound on the asymptotic parity-check density which gets the form

$$\liminf_{m\to\infty} \Delta_m \geq \frac{K_1 + K_2 \ln\left(\frac{1}{\varepsilon}\right)}{1-\varepsilon}$$

where

$$K_{1,2} = \frac{1-C}{C} \cdot K_{1,2}' \quad (2.32)$$

and $K'_{1,2}$ are introduced in (2.30) (note that $1 - R \geq 1 - C$). This completes the proof of the lower bound in (2.22) with the coefficients $K_{1,2}$ in (2.23).
Derivation of the optimization equation (2.24): We refer the reader to Appendix 2.A.2, where we also show the existence of such a solution. ■

Discussion: It is required to show that we achieve an improved lower bound on the parity-check density, as compared to the one in [81, Theorem 2.1]. To this end, it suffices to show that

$$\frac{(p_1 - p_2)^2}{p_1 + p_2} + \frac{(p_0 - p_3)^2}{p_0 + p_3} \geq (1 - 2w)^2. \tag{2.33}$$

For a proof of this inequality, we refer the reader to Appendix 2.A.3.

This therefore proves that the new lower bound is tighter (i.e., larger) than the original bound in [81, Theorem 2.1] (which corresponds to a two-level quantization of the LLR, as compared to the new bound which is based on a four-level quantization of the LLR).

Based on Proposition 2.1, we prove and discuss an upper bound on the asymptotic rate of every sequence of binary linear codes for which reliable communication is achievable. The bound refers to soft-decision ML decoding, and it is therefore valid for any suboptimal decoding algorithm. Hence, the following result also provides an upper bound on the achievable rate of ensembles of LDPC codes under iterative decoding where the transmission takes places over an MBIOS channel. The following bound improves the bounds stated in [17, Theorems 1 and 2]:

Corollary 2.1 ["Four-Level Quantization" Upper Bound on the Asymptotic Achievable Rates of Sequences of Binary Linear Block Codes] Let $\{C_m\}$ be a sequence of binary linear block codes whose codewords are transmitted with equal probability over an MBIOS channel, and suppose that the block length of this sequence of codes tends to infinity as $m \to \infty$. Let $\Gamma_{k,m}$ be the fraction of the parity-check nodes of degree k in an arbitrary representation of the code $\mathcal{C}_m$ by a bipartite graph which corresponds to a full-rank parity-check matrix. Then a necessary condition for this sequence to achieve vanishing bit error probability as $m \to \infty$ is that the asymptotic rate R of this sequence satisfies

$$\begin{aligned} R \;\leq\; 1 - \max\Bigg\{ (1-C) \cdot \Bigg(\sum_k \Bigg\{ \Gamma_k \sum_{t=0}^{k} \binom{k}{t} (p_1+p_2)^t (p_0+p_3)^{k-t} \\ \cdot h_2\Bigg(\frac{1 - \big(1 - \frac{2p_2}{p_1+p_2}\big)^t \big(1 - \frac{2p_3}{p_0+p_3}\big)^{k-t}}{2} \Bigg) \Bigg\} \Bigg)^{-1}, \\ \frac{2(p_2+p_3)}{1 - \sum_k \Gamma_k \Big(1 - 2(p_2+p_3)\Big)^k} \Bigg\} \end{aligned} \tag{2.34}$$

where p_0, p_1, p_2, p_3 are introduced in (2.9), and Γ_k and R are introduced in (2.1).

Proof: The first term in the maximization on the RHS of (2.34) follows from (2.10) in Proposition 2.1 and (2.25) in Lemma 2.2. It follows directly by combining both inequalities, and letting the bit error probability $P_{\rm b}$ tend to zero. The second term follows from the proof of [81, Corollary 3.1] which is based on the erasure decomposition Lemma [71]. ■

2.3.2 Extension of the Bounds to 2^d Quantization Levels

Following the method introduced in Section 2.3.1, we commence by deriving a lower bound on the conditional entropy of a transmitted codeword given the received sequence.

Proposition 2.2 Let $\mathcal{C}$ be a binary linear block code of length n and rate R. Let $\mathbf{X} = (X_1, \ldots, X_n)$ and $\mathbf{Y} = (Y_1, \ldots, Y_n)$ designate the transmitted codeword and received sequence, respectively, when the communication takes place over an MBIOS channel with conditional pdf $p_{Y|X}(\cdot|\cdot)$. For an arbitrary $d \geq 2$ and $0 < l_{2^{d-1}-1} < \ldots < l_1 < l_0 \triangleq \infty$, let us define the set of probabilities $\{p_s\}_{s=0}^{2^d-1}$ as follows:

$$p_s \triangleq \begin{cases} \Pr\{l_{s+1} < \mathrm{LLR}(Y) \leq l_s | X = 0\} & s = 0, \ldots, 2^{d-1} - 2 \\ \Pr\{0 < \mathrm{LLR}(Y) \leq l_{2^{d-1}-1} | X = 0\} & \\ +\frac{1}{2}\Pr\{\mathrm{LLR}(Y) = 0 | X = 0\} & s = 2^{d-1} - 1 \\ \Pr\{-l_{2^{d-1}-1} \leq \mathrm{LLR}(Y) < 0 \,|\, X = 0\} & \\ +\frac{1}{2}\Pr\{\mathrm{LLR}(Y) = 0 | X = 0\} & s = 2^{d-1} \\ \Pr\{-l_{2^d-(s+1)} \leq \mathrm{LLR}(Y) < -l_{2^d-s} | X = 0\} & s = 2^{d-1}+1, \ldots, 2^d - 1. \end{cases} \tag{2.35}$$

For an arbitrary full-rank parity-check matrix of the code $\mathcal{C}$, let Γ_k designate the fraction of the parity-checks involving k variables. Then, the conditional entropy of the transmitted codeword given the received sequence satisfies

$$\frac{H(\mathbf{X}|\mathbf{Y})}{n} \geq 1 - C - (1-R) \sum_k \Bigg\{ \Gamma_k \sum_{\substack{k_0, \ldots, k_{2^{d-1}-1} \\ \sum_i k_i = k}} \binom{k}{k_0, \ldots, k_{2^{d-1}-1}} \prod_{i=0}^{2^{d-1}-1} (p_i + p_{2^d-1-i})^{k_i} \cdot h_2\left(\frac{1}{2}\left[1 - \prod_{i=0}^{2^{d-1}-1}\left(1 - \frac{2p_{2^d-1-i}}{p_i + p_{2^d-1-i}}\right)^{k_i}\right]\right)\Bigg\}. \tag{2.36}$$

Proof: Following the proof of Proposition 2.1, we introduce a new physically degraded channel. It is a memoryless binary-input 2^d-ary output symmetric channel (see Figure 2.1 for $d = 2$). To this end, let $l_{2^{d-1}-1} < \ldots < l_1$ be arbitrary positive numbers, and denote $l_0 \triangleq \infty$. The output alphabet of the degraded channel is defined to be GF(2^d) whose elements form the set

$$\left\{ \sum_{j=0}^{d-1} a_j \, \alpha^j \; : \quad (a_0, a_1, \ldots, a_{d-1}) \in \{0,1\}^d \right\}$$

where α is a primitive element of GF(2^d).

For $s = 0, 1, \ldots, 2^{d-1} - 1$, denote the $(d-1)$–bit binary representation of s by $(a_1^{(s)}, \ldots, a_{d-1}^{(s)})$, i.e.,

$$s = \sum_{j=1}^{d-1} a_j^{(s)} \, 2^{j-1}.$$

Let X_i and Y_i designate the random variables referring to the input and output of the original channel $p_{Y|X}$ at time i (where $i = 1, \ldots, n$). As a natural generalization of the channel model in Figure 2.1, we introduce a physically degraded channel with 2^d quantization levels of the LLR. The output of this channel at time i, Z_i, is calculated from the output Y_i of the original channel as follows:

$$Z_i = \Phi_i + \Omega_i \tag{2.37}$$

where Φ_i is zero or one according to the sign of LLR(Y_i). It is set to zero or one, if the LLR is positive or negative, respectively; if LLR(Y_i) = 0, then Φ_i is either zero or one with equal probability. The value of Ω_i is calculated based on the absolute value of LLR(Y_i) as follows:

- If $l_{s+1} < |\mathrm{LLR}(Y_i)| \leq l_s$ for some $0 \leq s < 2^{d-1} - 1$, then

$$\Omega_i = \sum_{j=1}^{d-1} a_j^{(s)} \, \alpha^j. \tag{2.38}$$

- If $0 \leq |\mathrm{LLR}(Y_i)| \leq l_{2^{d-1}-1}$, then

$$\Omega_i = \sum_{j=1}^{d-1} \alpha^j. \tag{2.39}$$

From (2.35), the transition probabilities of the degraded channel are given by

$$\begin{aligned} p_s &= \Pr(Z = \sum_{j=1}^{d-1} a_j^{(s)} \, \alpha^j | X = 0) = \Pr(Z = 1 + \sum_{j=1}^{d-1} a_j^{(s)} \, \alpha^j | X = 1) \\ p_{2^d-1-s} &= \Pr(Z = 1 + \sum_{j=1}^{d-1} a_j^{(s)} \, \alpha^j | X = 0) = \Pr(Z = \sum_{j=1}^{d-1} a_j^{(s)} \, \alpha^j | X = 1) \end{aligned} \tag{2.40}$$

where $s = 0, 1, \ldots, 2^{d-1}-1$. The symmetry in these equalities holds since the channel is MBIOS.

Equations (2.11)–(2.15) hold also for the case of 2^d-level quantization. Thus, we will calculate the entropy of the random variable Z, and an upper bound on the entropy of the random vector $\mathbf{Z}$. This will finally provide the lower bound in (2.36).

Analogously to the proof of Proposition 2.1, the degraded channel is additive over GF(2^d). We denote the additive noise by

$$N_i = \Theta_i + \Omega_i. \tag{2.41}$$

Note that since the code is binary, then $\Phi_i = \Theta_i + X_i$, and the value of Ω_i stays the same in (2.37) and (2.41). Let $\mathbf{\Theta} \triangleq (\Theta_1, \ldots, \Theta_n)$ and $\mathbf{\Omega} \triangleq (\Omega_1, \ldots, \Omega_n)$. Due to the symmetry of the communication channel, it follows that $\mathbf{\Theta}$ and $\mathbf{\Omega}$ are statistically independent of the transmitted codeword $\mathbf{X}$. This gives

$$\begin{aligned} H(Z) &= H(\Phi, \Omega) \\ &= H(\Omega) + H(\Phi|\Omega) \\ &= H(\Omega) + 1 \end{aligned} \tag{2.42}$$

where the last equality follows from the same argument which validates (2.17).

We now derive an upper bound on the entropy of the random vector $\mathbf{Z}$. From the same chain of equalities leading to (2.18), it follows that

$$H(\mathbf{Z}) = nH(\Omega) + H(\mathbf{\Phi} \,|\, \mathbf{\Omega}) \tag{2.43}$$

where $\mathbf{\Phi} \triangleq (\Phi_1, \ldots, \Phi_n)$. As in the proof of Proposition 2.1, we define the syndrome as $\mathbf{S} = \mathbf{\Phi}\, H^T$ where H is a full-rank parity-check matrix of the code $\mathcal{C}$. As before, the calculation of the syndrome $\mathbf{S}$ only takes into account the Φ-components of the vector $\mathbf{Z}$. Since $\mathbf{X}\, H^T = 0$, then $\mathbf{S} = \mathbf{\Theta}\, H^T$ which is independent of the transmitted codeword. In parallel to (2.19), we obtain

$$H(\mathbf{\Phi} \,|\, \mathbf{\Omega}) \leq nR + H(\mathbf{S} \,|\, \mathbf{\Omega}). \tag{2.44}$$

Consider a parity-check equation which involves k variables, and let $\{i_1, \ldots, i_k\}$ be the set of indices of the variables involved in this parity-check equation. The component of the syndrome $\mathbf{S}$ which refers to this parity-check equation is zero if and only if the binary sub-vector $(\Theta_{i_1}, \ldots, \Theta_{i_k})$ has an even Hamming weight.

Lemma 2.4 Given that $(\Omega_{i_1}, \ldots, \Omega_{i_k})$ has k_s elements equal to $\sum_{j=1}^{d-1} a_j^{(s)}\, \alpha^j$ ($s = 0, \ldots, 2^{d-1} - 1$), the probability that the corresponding component of the syndrome

S_l is equal to 1 is given by

$$\frac{1}{2}\left[1-\prod_{s=0}^{2^{d-1}-1}\left(1-\frac{2p_{2^d-1-s}}{p_s+p_{2^d-1-s}}\right)^{k_s}\right].$$

Proof: From the probabilities which are associated with the quantized values of the LLR in (2.40), we get that for $0 \le s \le 2^{d-1}-1$

$$\Pr(\Theta_i = 1 \,|\, \Omega_i = \sum_{j=1}^{d-1} a_j^{(s)} \alpha^j) = \frac{p_{2^d-1-s}}{p_s + p_{2^d-1-s}}.$$

Since there are k_s indices i in the set $\{i_1, \ldots, i_k\}$ for which $\Omega_i = \sum_{j=1}^{d-1} a_j^{(s)} \alpha^j$, the lemma follows from [30, Lemma 4.1]. ■

Based on Lemma 2.4 and the discussion above, it follows that for any vector $\omega = (\omega_1, \ldots, \omega_k)$ which has k_s elements equal to $\sum_{j=1}^{d-1} a_j^{(s)} \alpha^j$ (where $s = 0, \ldots, 2^{d-1}-1$)

$$H\big(S_i \,|\, (\Omega_{i_1}, \ldots, \Omega_{i_k}) = \omega\big) = h_2\left(\frac{1}{2}\left[1-\prod_{s=0}^{2^{d-1}-1}\left(1-\frac{2p_{2^d-1-s}}{p_s+p_{2^d-1-s}}\right)^{k_s}\right]\right). \quad (2.45)$$

For a component S_i (where $1 \le i \le n(1-R)$) of the syndrome $\mathbf{S}$ which refers to a parity-check equation involving k variables

$$\begin{aligned}
&H(S_i \,|\, \mathbf{\Omega}) \\
&= H(S_i \,|\, \Omega_{i_1}, \ldots, \Omega_{i_k}) \\
&= \sum_{\omega} \Big\{ \Pr\big((\Omega_{i_1}, \ldots, \Omega_{i_k}) = \omega\big)\, H\big(S_i \,|\, (\Omega_{i_1}, \ldots, \Omega_{i_k}) = \omega\big) \Big\} \\
&= \sum_{\substack{k_0, \ldots, k_{2^{d-1}-1} \\ \sum_s k_s = k}} \Bigg\{ \binom{k}{k_0, \ldots, k_{2^{d-1}-1}} \prod_{s=0}^{2^{d-1}-1} (p_s + p_{2^d-1-s})^{k_s} \\
&\qquad\qquad \cdot h_2\left(\frac{1}{2}\left[1-\prod_{s=0}^{2^{d-1}-1}\left(1-\frac{2p_{2^d-1-s}}{p_s+p_{2^d-1-s}}\right)^{k_s}\right]\right)\Bigg\}
\end{aligned}$$

where the last equality follows since there are $\binom{k}{k_0, \ldots, k_{2^{d-1}-1}}$ vectors ω which have k_s elements of the type $\sum_{j=1}^{d-1} a_j^{(s)} \alpha^j$ for $s \in \{0, \ldots, 2^{d-1}-1\}$, and it also follows from the statistical independence of the components of the vector $\mathbf{\Omega}$, and from Eqs. (2.40) and (2.45).

The number of parity-check equations involving k variables is $n(1-R)\Gamma_k$, hence

$$\begin{aligned}
&H(\mathbf{S}\,|\,\mathbf{\Omega}) \\
&\leq \sum_{i=1}^{n(1-R)} H(S_i\,|\,\mathbf{\Omega}) \\
&= n(1-R)\sum_k \Bigg\{ \Gamma_k \sum_{\substack{k_0,\ldots,k_{2^{d-1}-1} \\ \sum_s k_s = k}} \binom{k}{k_0,\ldots,k_{2^{d-1}-1}} \prod_{s=0}^{2^{d-1}-1} (p_s + p_{2^d-1-s})^{k_s} \\
&\qquad\qquad \cdot h_2\left(\frac{1}{2}\left[1 - \prod_{s=0}^{2^{d-1}-1}\left(1 - \frac{2p_{2^d-1-s}}{p_s + p_{2^d-1-s}}\right)^{k_s}\right]\right)\Bigg\}. \qquad (2.46)
\end{aligned}$$

By combining (2.43)–(2.46), an upper bound on the entropy of the random vector $\mathbf{Z}$ follows:

$$\begin{aligned}
&H(\mathbf{Z}) \leq nR + nH(\Omega) \\
&\qquad + n(1-R)\sum_k \Bigg\{ \Gamma_k \sum_{\substack{k_0,\ldots,k_{2^{d-1}-1} \\ \sum_s k_s = k}} \binom{k}{k_0,\ldots,k_{2^{d-1}-1}} \prod_{s=0}^{2^{d-1}-1} (p_s + p_{2^d-1-s})^{k_s} \\
&\qquad\qquad \cdot h_2\left(\frac{1}{2}\left[1 - \prod_{s=0}^{2^{d-1}-1}\left(1 - \frac{2p_{2^d-1-s}}{p_s + p_{2^d-1-s}}\right)^{k_s}\right]\right)\Bigg\}. \qquad (2.47)
\end{aligned}$$

The substitution of (2.42) and (2.47) in (2.15) finally provides the lower bound on the conditional entropy $H(\mathbf{X}\,|\,\mathbf{Y})$ in (2.36). ■

Discussion: In the proof of Proposition 2.2, the upper bound on entropy of the degraded channel output $\mathbf{Z}$ is of the form

$$H(\mathbf{Z}) \leq nH(\Omega) + nR + \sum_{i=1}^{n(1-R)} H(S_i|\mathbf{\Omega})$$

which follows directly from (2.43), (2.44) and (2.46). Here, Ω is defined according to the conditions stated in (2.38) and (2.39), and S_i is the i^{th} component of the syndrome $\mathbf{S}$. Substituting (2.42) and the above inequality in (2.15) yields

$$\frac{H(X|Y)}{n} \geq 1 - C - \frac{1}{n}\sum_{i=1}^{n(1-R)} H(S_i|\mathbf{\Omega})$$

and the lower bound in (2.36) is in fact an explicit expression for the above inequality. The calculation of the lower bound in the RHS of (2.36) becomes more complex as the value of d is increased. However, when the quantization levels $l_1, \ldots, l_{2^{d-1}-1}$ are

set to maximize the lower bound in the RHS of (2.36), we show that the bound is monotonically increasing with d.

To this end, let $d \geq 2$ be an arbitrary integer, $(l_1^{(d)}, \ldots, l_{2^{d-1}-1}^{(d)})$ with their symmetric values around zero be the optimal choice of 2^d quantization levels, and denote the random variable Ω for this setting by $\Omega^{(d)}$ (see (2.38) and (2.39)). Consider any set of 2^{d+1} quantization levels $(l_1^{(d+1)}, \ldots, l_{2d-1}^{(d+1)})$ with their symmetric values around zero such that $l_{2i}^{(d+1)} = l_i^{(d)}$ for $i = 1, \ldots, 2^{d-1} - 1$. Denote the random variable Ω for this choice of quantization levels by $\Omega^{(d+1)}$. Clearly, since the former set of 2^d quantization levels is a subset of the latter set of 2^{d+1} levels, then $\Omega^{(d)}$ can be calculated from $\Omega^{(d+1)}$. Let $\mathbf{\Omega}^{(k)} \triangleq (\Omega_1^{(k)}, \ldots, \Omega_n^{(k)})$ for $k = d$ and $d+1$. By the information processing inequality

$$1 - C - \frac{1}{n} \sum_{i=1}^{n(1-R)} H(S_i|\mathbf{\Omega}^{(d)}) \leq 1 - C - \frac{1}{n} \sum_{i=1}^{n(1-R)} H(S_i|\mathbf{\Omega}^{(d+1)}).$$

Therefore, the (possibly sub-optimal) set of 2^{d+1} quantization levels $(l_1^{(d+1)}, \ldots, l_{2d-1}^{(d+1)})$ with their symmetric values around zero provides a tighter lower bound than the *optimal* choice of 2^d quantization levels. Hence, this proves that the lower bound is monotonically increasing with the number of quantization levels when these levels are set optimally.

Theorem 2.4 ["2^d-Level Quantization" Lower Bound on the Asymptotic Parity-Check Density of Binary Linear Block Codes] Let $\{C_m\}$ be a sequence of binary linear block codes achieving a fraction $1 - \varepsilon$ of the capacity of an MBIOS channel with vanishing bit error probability. Let H_m be an arbitrary *full-rank* parity-check matrix of the code $\mathcal{C}_m$, and denote its density by Δ_m. Then, the asymptotic density satisfies

$$\liminf_{m \to \infty} \Delta_m > \frac{K_1 + K_2 \ln \frac{1}{\varepsilon}}{1 - \varepsilon} \tag{2.48}$$

where

$$K_1 = K_2 \ln \left(\frac{1}{2 \ln 2} \frac{1 - C}{C} \right), \quad K_2 = -\frac{1 - C}{C \ln \left(\sum_{i=0}^{2^{d-1}-1} \frac{(p_i - p_{2^d-1-i})^2}{p_i + p_{2^d-1-i}} \right)}. \tag{2.49}$$

Here, $d \geq 2$ is an arbitrary integer and the probabilities $\{p_i\}$ are introduced in (2.35) in terms of $l_1 > \ldots > l_{2^{d-1}-1} \in \mathbb{R}^+$. The optimal vector of quantization levels $(l_1, \ldots, l_{2^{d-1}-1})$ is given implicitly by solving the set of $2^{d-1} - 1$ equations

$$\frac{p_{2^d-1-i}^2 + e^{-l_i} p_i^2}{(p_i + p_{2^d-1-i})^2} = \frac{p_{2^d-i}^2 + e^{-l_i} p_{i-1}^2}{(p_{i-1} + p_{2^d-i})^2}, \quad i = 1, \ldots, 2^{d-1} - 1. \tag{2.50}$$

where such a solution always exists.[3]

Proof: For an arbitrary sequence of binary linear block codes $\{\mathcal{C}_m\}$ which achieves a fraction $1-\varepsilon$ to capacity with vanishing bit error probability, we get from Lemma 2.2 that

$$\lim_{m\to\infty} \frac{H(\mathbf{X_m} \,|\, \mathbf{Y_m})}{n_m} = 0$$

where $\mathbf{X_m}$ and $\mathbf{Y_m}$ designate the transmitted codeword in the code $\mathcal{C}_m$ and the received sequence, respectively, and n_m designates the block length of the code $\mathcal{C}_m$. From Proposition 2.2, we obtain

$$\begin{aligned}\frac{H(\mathbf{X_m}|\mathbf{Y_m})}{n_m} \geq & 1 - C - (1 - R_m) \\ & \cdot \sum_k \left\{ \Gamma_{k,m} \sum_{\substack{k_0,\ldots,k_{2^{d-1}-1} \\ \sum_s k_s = k}} \binom{k}{k_0, \ldots, k_{2^{d-1}-1}} \prod_{s=0}^{2^{d-1}-1} (p_s + p_{2^d-1-s})^{k_s} \right. \\ & \left. \cdot h_2 \left(\frac{1}{2} \left[1 - \prod_{s=0}^{2^{d-1}-1} \left(1 - \frac{2p_{2^d-1-s}}{p_s + p_{2^d-1-s}} \right)^{k_s} \right] \right) \right\}.\end{aligned}$$

where $\Gamma_{k,m}$ designates the fraction of parity-check equations in a parity-check matrix H_m which involve k variables. The upper bound on the binary entropy function h_2 in Lemma 2.3 gives

$$\begin{aligned}1 - C - \big(1 - R_m\big) \; & \sum_k \left\{ \Gamma_{k,m} \sum_{\substack{k_0,\ldots,k_{2^{d-1}-1} \\ \sum_s k_s = k}} \binom{k}{k_0, \ldots, k_{2^{d-1}-1}} \prod_{s=0}^{2^{d-1}-1} (p_s + p_{2^d-1-s})^{k_s} \right. \\ & \left. \left[1 - \frac{1}{2\ln 2} \left(\prod_{s=0}^{2^{d-1}-1} \frac{p_s - p_{2^d-1-s}}{p_s + p_{2^d-1-s}} \right)^{2k_s} \right] \right\} \leq \frac{H(\mathbf{X_m}|\mathbf{Y_m})}{n_m}. \quad (2.51)\end{aligned}$$

Since $\sum_k \Gamma_{k,m} = 1$ and $\sum_{s=0}^{2^d-1} p_s = 1$, we get

$$\begin{aligned}& 1 - \frac{1}{2\ln 2} \sum_k \left\{ \Gamma_{k,m} \sum_{\substack{k_0,\ldots,k_{2^{d-1}-1} \\ \sum_s k_s = k}} \binom{k}{k_0, \ldots, k_{2^{d-1}-1}} \prod_{s=0}^{2^{d-1}-1} \left(\frac{(p_s - p_{2^d-1-s})^2}{p_s + p_{2^d-1-s}} \right)^{k_s} \right\} \\ & = 1 - \frac{1}{2\ln 2} \sum_k \left\{ \Gamma_{k,m} \left(\sum_{s=0}^{2^{d-1}-1} \frac{(p_s - p_{2^d-1-s})^2}{p_s + p_{2^d-1-s}} \right)^k \right\}\end{aligned}$$

[3] See the footnote to Theorem 2.3 on p. 33.

$$\leq 1 - \frac{1}{2\ln 2} \left(\sum_{s=0}^{2^{d-1}-1} \frac{(p_s - p_{2^d-1-s})^2}{p_s + p_{2^d-1-s}} \right)^{a_{\mathrm{R}}(m)} \tag{2.52}$$

where $a_{\mathrm{R}}(m) \triangleq \sum_k k\Gamma_{k,m}$ designates the average right degree of the bipartite graph which refers to the parity-check matrix H_m, and the last transition follows from Jensen's inequality.

Substituting the RHS of (2.52) into the LHS of (2.51) and letting m tend to infinity gives the inequality

$$1 - C - \big(1 - (1-\varepsilon)C\big) \left[1 - \frac{1}{2\ln 2} \left(\sum_{s=0}^{2^{d-1}-1} \frac{(p_s - p_{2^d-1-s})^2}{p_s + p_{2^d-1-s}} \right)^{a_{\mathrm{R}}(\infty)} \right] \leq 0 \tag{2.53}$$

where $a_{\mathrm{R}}(\infty) \triangleq \liminf_{m\to\infty} a_{\mathrm{R}}(m)$. Note that the validity of (2.53) follows since the base of the exponent in the inequality above does not exceed unity, i.e.,

$$\begin{aligned} &\sum_{s=0}^{2^{d-1}-1} \frac{(p_s - p_{2^d-1-s})^2}{p_s + p_{2^d-1-s}} \\ &\leq \sum_{s=0}^{2^{d-1}-1} \frac{(p_s + p_{2^d-1-s})^2}{p_s + p_{2^d-1-s}} \\ &= \sum_{s=0}^{2^{d-1}-1} p_s + p_{2^d-1-s} = 1. \end{aligned}$$

Inequality (2.53) gives the following lower bound on the asymptotic average right degree:

$$a_{\mathrm{R}} \geq K_1' + K_2' \ln\left(\frac{1}{\varepsilon}\right) \tag{2.54}$$

where

$$K_1' = -\frac{\ln\left(\frac{1}{2\ln 2}\frac{1-C}{C}\right)}{\ln\left(\sum_{s=0}^{2^{d-1}-1} \frac{(p_s - p_{2^d-1-s})^2}{p_s + p_{2^d-1-s}} \right)}, \quad K_2' = -\frac{1}{\ln\left(\sum_{s=0}^{2^{d-1}-1} \frac{(p_s - p_{2^d-1-s})^2}{p_s + p_{2^d-1-s}} \right)}.$$

By combining (2.31) and (2.54) with the asymptotic rate $R = (1-\varepsilon)C$, we obtain a lower bound on the asymptotic parity-check density which is of the form

$$\liminf_{m\to\infty} \Delta_m \geq \frac{K_1 + K_2 \ln\left(\frac{1}{\varepsilon}\right)}{1-\varepsilon}$$

where

$$K_{1,2} = \frac{1-C}{C} \cdot K_{1,2}'.$$

This completes the proof of the lower bound in (2.48) and (2.49). The derivation of the set of optimization equations in (2.50) follows along the lines of the derivation of (2.24). In the general case of 2^d quantization levels, it follows from (2.49) that we need to maximize

$$\sum_{s=0}^{2^{d-1}-1} \frac{(p_s - p_{2^d-1-s})^2}{p_s + p_{2^d-1-s}}.$$

To this end, we set to zero all the partial derivatives w.r.t. l_s where $s = 1, \ldots, 2^{d-1}-1$. Since from (2.35) only p_s, p_{s-1}, p_{2^d-s} and p_{2^d-s-1} depend on l_s, then

$$\frac{\partial}{\partial l_s}\left\{\frac{(p_{s-1} - p_{2^d-s})^2}{p_{s-1} + p_{2^d-s}} + \frac{(p_s - p_{2^d-s-1})^2}{p_s + p_{2^d-s-1}}\right\} = 0.$$

We express now the probabilities p_s, p_{s-1}, p_{2^d-s} and p_{2^d-s-1} as integrals of the conditional pdf a of the LLR, and rely on the symmetry property which states that $a(l) = e^l a(-l)$ for $l \in \mathbb{R}$. In a similar manner to the derivation of (2.24), this gives the set of equations in (2.50). Their solution provides the quantization levels $l_1, \ldots, l_{2^{d-1}-1}$ (where according to Proposition 2.3.2, the other $2^{d-1} - 1$ levels are set to be symmetric w.r.t. zero). ■

Based on the proof of Theorem 2.4, we derive an upper bound on the asymptotic rate of every sequence of binary linear codes for which reliable communication is achievable. The bound refers of soft-decision ML decoding, and it is therefore valid for any sub-optimal decoding algorithm.

Corollary 2.2 ["2^d-Level Quantization" Upper Bound on the Asymptotic Achievable Rates of Sequences of Binary Linear Block Codes] Let $\{C_m\}$ be a sequence of binary linear block codes whose codewords are transmitted with equal probability over an MBIOS channel, and suppose that the block length of this sequence of codes tends to infinity as $m \to \infty$. Let $\Gamma_{k,m}$ be the fraction of the parity-check nodes of degree k in an arbitrary representation of the code $\mathcal{C}_m$ by a bipartite graph which corresponds to a full-rank parity-check matrix. Then a necessary condition for this sequence to achieve vanishing bit error probability as $m \to \infty$ is that the asymptotic rate R of this sequence satisfies

$$R \leq 1 - \max\left\{(1-C)\left\{\sum_k \Gamma_k \sum_{\substack{k_0,\ldots,k_{2^{d-1}-1} \\ \cdot \sum_i k_i = k}} \binom{k}{k_0, \ldots, k_{2^{d-1}-1}} \prod_{i=0}^{2^{d-1}-1} (p_i + p_{2^d-1-i})^{k_i} \right.\right.$$
$$\left.\left. \cdot\, h_2\left(\frac{1}{2}\left[1 - \prod_{i=0}^{2^{d-1}-1}\left(1 - \frac{2p_{2^d-1-i}}{p_i + p_{2^d-1-i}}\right)^{k_i}\right]\right)\right\}^{-1}, \right.$$
$$\left. \frac{2\sum_{i=2^{d-1}}^{2^d-1} p_i}{1 - \sum_k \Gamma_k \left(1 - 2\sum_{i=2^{d-1}}^{2^d-1} p_i\right)^k}\right\} \tag{2.55}$$

where $d \geq 2$ is arbitrary, the probabilities $\{p_i\}$ are introduced in (2.35), and Γ_k and R are introduced in (2.1).

Proof: The concept of the proof is the same as the proof of Corollary 2.1, except that the first term in the RHS of (2.55) relies on (2.36). ■

2.4 Approach II: Bounds without Quantization of the LLR

Similarly to the previous section, we derive bounds on the asymptotic achievable rate and the asymptotic parity-check density of an arbitrary sequence of binary linear block codes transmitted over an MBIOS channel. As in Section 2.3, the derivation of these two bounds is based on a lower bound on the conditional entropy of a transmitted codeword given the received sequence at the output of an arbitrary MBIOS channel.

Proposition 2.3 Let $\mathcal{C}$ be a binary linear block code of length n and rate R transmitted over an MBIOS channel. Let $\mathbf{X} = (X_1, \ldots, X_n)$ and $\mathbf{Y} = (Y_1, \ldots, Y_n)$ designate the transmitted codeword and the received sequence, respectively. For an arbitrary representation of the code $\mathcal{C}$ by a full-rank parity-check matrix, let Γ_k designate the fraction of the parity-check equations of degree k, and $\Gamma(x) \triangleq \sum_k \Gamma_k x^k$ be the degree distribution of the parity-check nodes in the corresponding bipartite graph. Then, the conditional entropy of the transmitted codeword given the received sequence satisfies

$$\frac{H(\mathbf{X}|\mathbf{Y})}{n} \geq 1 - C - (1-R)\left(1 - \frac{1}{2\ln 2}\sum_{p=1}^{\infty}\frac{\Gamma(g_p)}{p(2p-1)}\right) \tag{2.56}$$

where

$$g_p \triangleq \int_0^{\infty} a(l)(1+e^{-l})\tanh^{2p}\left(\frac{l}{2}\right)\mathrm{d}l\,, \quad p \in \mathbb{N} \tag{2.57}$$

and a denotes the conditional pdf of the LLR given that the channel input is 0.

Proof: We consider a binary linear block code $\mathcal{C}$ of length n and rate R whose transmission takes place over an MBIOS channel. For the continuation of the proof, we move from the mapping of the MBIOS channel $X \to Y$ to the channel $X \to \widetilde{Y}$ where $\widetilde{Y}$ represents the LLR of the channel output Y. These channels are equivalent in the sense that $H(X|Y) = H(X|\widetilde{Y})$. The basic idea for showing the equivalence between the original channel and the one which will be introduced shortly is based on the fact that the LLR forms a sufficient statistics of the channel.

For the characterization of the equivalent channel, let the function a designate the conditional pdf of the LLR given that the channel input is 0. We randomly generate an i.i.d. sequence $\{L_i\}_{i=1}^n$ w.r.t. the conditional pdf a, and define

$$\Omega_i \triangleq |L_i|, \quad \Theta_i \triangleq \begin{cases} 0 & \text{if} \quad L_i > 0 \\ 1 & \text{if} \quad L_i < 0 \\ 0 \text{ or } 1 \text{ w.p. } \frac{1}{2} & \text{if} \quad L_i = 0 \end{cases} . \tag{2.58}$$

The output of the equivalent channel is defined to be the sequence $\widetilde{\mathbf{Y}} = (\widetilde{Y}_1, \ldots, \widetilde{Y}_n)$ where

$$\widetilde{Y}_i = (\Phi_i, \Omega_i), \quad i = 1, \ldots, n$$

and $\Phi_i = \Theta_i + X_i$ where this addition is modulo-2. The output of this equivalent channel at time i is therefore the pair (Φ_i, Ω_i) where $\Phi_i \in \{0, 1\}$ and $\Omega_i \in \mathbb{R}^+$. This defines the memoryless mapping

$$X \to \widetilde{Y} \triangleq (\Phi, \Omega)$$

where Φ is a binary random variable which is affected by X, and Ω is a non-negative random variable which is not affected by X. Note that due to the symmetry of the communication channel, the joint distribution of the pair (Φ, Ω) is equal to the one which corresponds to the pair representing the sign and magnitude of LLR(Y). Hence,

$$f_\Omega(\omega) = \begin{cases} a(\omega) + a(-\omega) = (1 + e^{-\omega})\, a(\omega) & \text{if} \quad \omega > 0 \\ a(0) & \text{if} \quad \omega = 0 \end{cases} \tag{2.59}$$

where we rely on the symmetry property of the pdf a.

Following the lines which lead to (2.15), we obtain

$$H(\mathbf{X}|\mathbf{Y}) \geq nR - H(\widetilde{\mathbf{Y}}) + nH(\widetilde{Y}) - nC. \tag{2.60}$$

In order to get a lower bound on $H(\mathbf{X}|\mathbf{Y})$, we calculate the entropy of $\widetilde{Y}$ and also obtain an upper bound on the entropy of $\widetilde{\mathbf{Y}}$. The calculation of the first entropy is direct

$$\begin{aligned} H(\widetilde{Y}) &= H(\Phi, \Omega) \\ &= H(\Omega) + H(\Phi|\Omega) \\ &= H(\Omega) + E_\omega \left[H(\Phi|\Omega = \omega) \right] \\ &= H(\Omega) + 1 \end{aligned} \tag{2.61}$$

where the last transition is due to the fact that given the absolute value of the LLR, its sign is equally likely to be positive or negative. The entropy $H(\Omega)$ is not expressed explicitly as it will cancel out later.

We now derive an upper bound on $H(\widetilde{\mathbf{Y}})$.

$$\begin{aligned} H(\widetilde{\mathbf{Y}}) &= H\big(\Phi_1, \Omega_1, \ldots, \Phi_n, \Omega_n\big) \\ &= H(\Omega_1, \ldots, \Omega_n) + H\big(\Phi_1, \ldots, \Phi_n \,|\, \Omega_1, \ldots, \Omega_n\big) \\ &= nH(\Omega) + H\big(\Phi_1, \ldots, \Phi_n \,|\, \Omega_1, \ldots, \Omega_n\big). \end{aligned} \tag{2.62}$$

Define the syndrome vector

$$\mathbf{S} = (\Phi_1, \ldots, \Phi_n)\, H^T$$

where H is an arbitrary full-rank parity-check matrix of the binary linear block code $\mathcal{C}$, and let M be the index of the vector $(\Phi_1, \ldots, \Phi_n)$ in the coset which corresponds to $\mathbf{S}$. Since each coset has exactly 2^{nR} elements which are equally likely then $H(M) = nR$, and we get

$$\begin{aligned} H\big((\Phi_1, \ldots, \Phi_n) \,|\, (\Omega_1, \ldots, \Omega_n)\big) &= H(\mathbf{S}, M \,|\, (\Omega_1, \ldots, \Omega_n)\big) \\ &\leq H(M) + H\big(\mathbf{S} \,|\, (\Omega_1, \ldots, \Omega_n)\big) \\ &= nR + H\big(\mathbf{S} \,|\, (\Omega_1, \ldots, \Omega_n)\big) \\ &\leq nR + \sum_{j=1}^{n(1-R)} H\big(S_j \,|\, (\Omega_1, \ldots, \Omega_n)\big) \end{aligned} \tag{2.63}$$

Since $\mathbf{X}\, H^T = \mathbf{0}$ for any codeword $\mathbf{X}$, then

$$\mathbf{S} = (\Theta_1, \ldots, \Theta_n)\, H^T.$$

which is independent of the transmitted codeword. Consider the j^{th} parity-check equation, and assume that it involves k variables whose indices are $i_1, \ldots, i_k$. Then, the component S_j of the syndrome is equal to 1 if and only if there is an odd number of ones in the random vector $(\Theta_{i_1}, \ldots, \Theta_{i_k})$.

Lemma 2.5 If the j-th component of the syndrome $\mathbf{S}$ involves k variables whose indices are $i_1, i_2, \ldots, i_k$, then

$$\Pr\big(S_j = 1 \,|\, (\Omega_{i_1}, \ldots, \Omega_{i_k}) = (\alpha_1, \ldots, \alpha_k)\big) = \frac{1}{2}\left[1 - \prod_{m=1}^{k} \tanh\left(\frac{\alpha_m}{2}\right)\right]. \tag{2.64}$$

Proof: Due to the symmetry of the channel

$$\begin{aligned} P_m &\triangleq \Pr(\Theta_{i_m} = 1 \,|\, \Omega_{i_m} = \alpha_m) \\ &= \frac{a(-\alpha_m)}{a(\alpha_m) + a(-\alpha_m)} \\ &= \frac{1}{1 + e^{\alpha_m}} \end{aligned}$$

Substituting this result in [30, Lemma 4.1] gives

$$\begin{aligned}
&\Pr\big(S_j = 1 \,|\, (\Omega_{i_1}, \ldots, \Omega_{i_k}) = (\alpha_1, \ldots, \alpha_k)\big) \\
&= \frac{1}{2}\left[1 - \prod_{m=1}^{k}(1 - 2P_m)\right] \\
&= \frac{1}{2}\left[1 - \prod_{m=1}^{k}\tanh\left(\frac{\alpha_m}{2}\right)\right], \quad m = 1, \ldots, k.
\end{aligned}$$

■

We therefore obtain from Lemma 2.5 that

$$H\big(S_j|(\Omega_{i_1}, \ldots, \Omega_{i_k}) = (\alpha_1, \ldots, \alpha_k)\big) = h_2\left(\frac{1}{2}\left[1 - \prod_{m=1}^{k}\tanh\left(\frac{\alpha_m}{2}\right)\right]\right)$$

and by taking the statistical expectation over the k random variables $\Omega_{i_1}, \ldots, \Omega_{i_k}$, we get

$$\begin{aligned}
&H\big(S_j|\Omega_{i_1}, \ldots, \Omega_{i_k}\big) \\
&= \int_0^\infty \ldots \int_0^\infty h_2\left(\frac{1}{2}\left[1 - \prod_{m=1}^{k}\tanh\left(\frac{\alpha_m}{2}\right)\right]\right)\prod_{m=1}^{k} f_\Omega(\alpha_m)\,\mathrm{d}\alpha_1\mathrm{d}\alpha_2\ldots\mathrm{d}\alpha_k \qquad (2.65)\\
&= 1 - \frac{1}{2\ln 2}\sum_{p=1}^{\infty}\left\{\frac{1}{p(2p-1)}\left(\int_0^\infty f_\Omega(\alpha)\,\tanh^{2p}\left(\frac{\alpha}{2}\right)\,\mathrm{d}\alpha\right)^k\right\}
\end{aligned}$$

where the equality in the last transition is proved in Appendix 2.B.2. Hence, since $n(1-R)\Gamma_k$ designates the number of parity-check equations of degree k, then

$$\begin{aligned}
&\sum_{j=1}^{n(1-R)} H\big(S_j|\Omega_1, \ldots, \Omega_n\big) \\
&= n(1-R)\left[1 - \frac{1}{2\ln 2}\sum_k\left\{\Gamma_k\sum_{p=1}^{\infty}\frac{1}{p(2p-1)}\left(\int_0^\infty f_\Omega(\alpha)\,\tanh^{2p}\left(\frac{\alpha}{2}\right)\,\mathrm{d}\alpha\right)^k\right\}\right] \\
&= n(1-R)\left[1 - \frac{1}{2\ln 2}\sum_k\left\{\Gamma_k\sum_{p=1}^{\infty}\frac{g_p^k}{p(2p-1)}\right\}\right] \qquad (2.66)
\end{aligned}$$

where the last equality follows from (2.57) and (2.59). By combining (2.62), (2.63) and (2.66), we get the following upper bound on $H(\widetilde{\mathbf{Y}})$:

$$\begin{aligned}
H(\widetilde{\mathbf{Y}}) &\le nH(\Omega) + nR + n(1-R)\left[1 - \frac{1}{2\ln 2}\sum_k\left\{\Gamma_k\sum_{p=1}^{\infty}\frac{g_p^k}{p(2p-1)}\right\}\right] \\
&= nH(\Omega) + nR + n(1-R)\left[1 - \frac{1}{2\ln 2}\sum_{p=1}^{\infty}\left\{\frac{1}{p(2p-1)}\sum_k\Gamma_k g_p^k\right\}\right] \\
&= nH(\Omega) + nR + n(1-R)\left(1 - \frac{1}{2\ln 2}\sum_{p=1}^{\infty}\frac{\Gamma(g_p)}{p(2p-1)}\right). \qquad (2.67)
\end{aligned}$$

Finally, the equality in (2.61) and the upper bound on $H(\widetilde{\mathbf{Y}})$ given in (2.67) are substituted in the RHS of (2.60). This provides the lower bound on the conditional entropy $H(\mathbf{X}|\mathbf{Y})$ given in (2.56), and completes the proof of this proposition. ■

Remark 2.2 For the particular case of a BEC with erasure probability p, the capacity is $C = 1 - p$ bits per cannel use. The conditional pdf of the LLR, given that the 0 is transmitted, is equal to

$$a(l) = p\Delta_0(l) + (1-p)\Delta_\infty(l)$$

where the function Δ_a designates the Dirac Delta function at the point a. We obtain from (2.57) that $g_m = 1 - p$ for all $m \in \mathbb{N}$, so (2.56) gives

$$\begin{aligned} \frac{H(\mathbf{X}|\mathbf{Y})}{n} &\geq p - (1-R)\left[1 - \frac{1}{2\ln 2}\sum_{m=1}^{\infty}\frac{\Gamma(1-p)}{m(2m-1)}\right] \\ &= p - (1-R)\Big[1 - \Gamma(1-p)\Big] \end{aligned} \tag{2.68}$$

where this equality above follows since $\displaystyle\sum_{m=1}^{\infty}\frac{1}{2m(2m-1)} = \ln 2$.

This lower bound on the conditional entropy for the BEC coincides with the result proved in [81, Eqs. (33) and (34)]. The result there was obtained by the derivation of an upper bound on the rank of H_{E} which is a sub-matrix of H whose columns correspond to the variables erased by the BEC.

Discussion: The proof of Proposition 2.3 relies on the analysis of an equivalent channel rather than a degraded (quantized) channel. We therefore expect the lower bound in the RHS of (2.56) to be tighter than the one in the RHS of (2.36). By following the derivation in (2.60)–(2.63) gives

$$\frac{H(\mathbf{X}|\mathbf{Y})}{n} \geq 1 - C - \frac{1}{n}\sum_{j=1}^{n(1-R)} H(S_j|\Omega_1,\ldots,\Omega_n) \tag{2.69}$$

where the random variables $\Omega_1,\ldots,\Omega_n$ are defined in (2.58) and S_j is the j-th component of the syndrome. The lower bound in (2.56) is in fact an explicit expression for the above inequality where the side information Ω is the absolute value of the LLR without quantization. From the discussion following Proposition 2.2, the bound in (2.36) is of the same form, except that the side-information $\Omega_1,\ldots,\Omega_n$ is a *quantized version* of the absolute value of the LLR. Hence, from the information processing inequality it follows the indeed (2.56) is a tighter lower bound on the conditional entropy than (2.36) for any number of quantization levels.

Theorem 2.5 ["Un-Quantized" Lower Bound on the Asymptotic Parity-Check Density of Binary Linear Block Codes] Let $\{C_m\}$ be a sequence of binary linear codes achieving a fraction $1-\varepsilon$ of the capacity C of an MBIOS channel with vanishing bit error probability. Let H_m be an arbitrary *full-rank* parity-check matrix of the code $\mathcal{C}_m$, and denote its density by Δ_m. Then, the asymptotic density satisfies

$$\liminf_{m\to\infty} \Delta_m \geq \frac{K_1 + K_2 \ln \frac{1}{\varepsilon}}{1-\varepsilon} \tag{2.70}$$

where

$$K_1 = K_2 \ln\left(\frac{\xi\,(1-C)}{C}\right), \quad K_2 = \frac{1-C}{C}\,\frac{1}{\ln\left(\frac{1}{g_1}\right)}, \tag{2.71}$$

g_1 is introduced in (2.57), and

$$\xi \triangleq \begin{cases} 1 & \text{for a BEC} \\ \frac{1}{2\ln 2} & \text{otherwise} \end{cases}. \tag{2.72}$$

Proof: From the lower bound on $\frac{H(\mathbf{X}\,|\,\mathbf{Y})}{n}$ in Eq. (2.56) and Lemma 2.2 (see p. 33), we obtain that if $\{\mathcal{C}_m\}$ is a sequence of binary linear block codes which achieves a fraction $1-\varepsilon$ of the channel capacity with vanishing bit error probability, then

$$1 - C - \big(1-(1-\varepsilon)C\big)\left(1 - \frac{1}{2\ln 2}\sum_{p=1}^{\infty}\frac{\Gamma(g_p)}{p(2p-1)}\right) \leq 0$$

where $\Gamma(x) \triangleq \sum_k \Gamma_k x^k$ is the asymptotic right degree distribution from the node perspective which corresponds to the sequence of parity-check matrices $\{H_m\}$. Since $\sum_k k\Gamma_k = a_{\mathrm{R}}$ is the average right degree, then from the convexity of the exponential function, we obtain by invoking Jensen's inequality that

$$1 - C - \big(1-(1-\varepsilon)C\big)\left[1 - \frac{1}{2\ln 2}\sum_{p=1}^{\infty}\frac{g_p^{a_{\mathrm{R}}}}{p(2p-1)}\right] \leq 0. \tag{2.73}$$

We derive now two different lower bounds on the infinite sum in the LHS of (2.73), and compare them later. For the derivation of the lower bound in the first approach, let us define the positive sequence

$$\alpha_p \triangleq \frac{1}{\ln 2}\,\frac{1}{2p(2p-1)}, \quad p = 1, 2, \ldots \tag{2.74}$$

From (2.B.1) in the appendix, the substitution of $x = 0$ in both sides of the equality gives that $\sum_{p=1}^{\infty}\alpha_p = 1$, so the sequence $\{\alpha_p\}$ forms a probability distribution. We

therefore obtain that

$$\begin{aligned}
&\frac{1}{2\ln 2}\sum_{p=1}^{\infty}\frac{g_p^{a_{\mathrm{R}}}}{p(2p-1)} \\
&\overset{(a)}{=} \sum_{p=1}^{\infty}\left\{\alpha_p\left(\int_0^{\infty} a(l)(1+e^{-l})\tanh^{2p}\left(\frac{l}{2}\right)\,\mathrm{d}l\right)^{a_{\mathrm{R}}}\right\} \\
&\overset{(b)}{\geq} \left(\int_0^{\infty} a(l)(1+e^{-l})\sum_{p=1}^{\infty}\alpha_p\tanh^{2p}\left(\frac{l}{2}\right)\,\mathrm{d}l\right)^{a_{\mathrm{R}}} \\
&\overset{(c)}{=} \left(\int_0^{\infty} a(l)(1+e^{-l})\left[1-h_2\left(\frac{1}{2}\left[1-\tanh\left(\frac{l}{2}\right)\right]\right)\right]\mathrm{d}l\right)^{a_{\mathrm{R}}} \\
&\overset{(d)}{=} \left(\int_0^{\infty} a(l)(1+e^{-l})\left[1-h_2\left(\frac{1}{1+e^{l}}\right)\right]\,\mathrm{d}l\right)^{a_{\mathrm{R}}} \\
&\overset{(e)}{=} C^{a_R}
\end{aligned} \tag{2.75}$$

where equality (a) follows from (2.57) and (2.74), inequality (b) follows from Jensen's inequality, equality (c) follows from (2.74) and (2.B.1), equality (d) follows from the identity $\tanh(x) = \frac{e^{2x}-1}{e^{2x}+1}$, and equality (e) follows from the relation between the capacity of an MBIOS channel and the pdf of the absolute value of the LLR (see [74, Lemma 3.13]).

For an alternative derivation of the lower bound of the infinite series, we can truncate the infinite sum in the RHS of (2.73) and take into account only the first term in this series. This gives

$$\frac{1}{2\ln 2}\sum_{p=1}^{\infty}\frac{g_p^{a_{\mathrm{R}}}}{p(2p-1)} \geq \frac{g_1^{a_{\mathrm{R}}}}{2\ln 2} \tag{2.76}$$

which follows from (2.57) since $g_p \geq 0$ for all $p \in \mathbb{N}$.

In order to compare the tightness of the two lower bounds in (2.75) and (2.76), we first compare the bases of their exponents (i.e., g_1 and C). To this end, it is easy to verify that

$$\tanh^2\left(\frac{l}{2}\right) \geq 1 - h_2\left(\frac{1}{1+e^{l}}\right) \quad l \in [0,\infty)$$

with an equality if and only if $l = 0$ or $l \to \infty$. To show this, we start from equality (2.B.1), use the inequality $(1-2x)^{2p} \leq (1-2x)^2$ for $p \in \mathbb{N}$ and $0 \leq x \leq 1$, and the equality $\displaystyle\sum_{p=1}^{\infty}\frac{1}{2p(2p-1)} = \ln 2$ to finally get

$$h_2(x) \geq 1 - (1-2x)^2\,, \quad 0 \leq x \leq 1.$$

Hence, from (2.75) and (2.76), this gives $g_1 \geq C$ with equality if and only if the MBIOS channel is a BEC. Therefore, up to the multiplicative constant $\frac{1}{2\ln 2}$, the second lower bound is tighter than the first one. However, we note that for the BEC, the first bound is tighter. It gives an improvement by a factor of $2\ln 2 \approx 1.386$.

We will therefore continue the analysis based on the second bound in (2.76), and then give the potential improvement which follows from the first bound in (2.75) for a BEC. From (2.73) and (2.76), we obtain that

$$1 - C - \big(1 - (1-\varepsilon)C\big)\left(1 - \frac{g_1^{a_{\mathrm{R}}}}{2\ln 2}\right) \leq 0.$$

Since $g_1 \leq 1$ (where equality is achieved for a noiseless channel), then the asymptotic average right degree (a_{R}) satisfies the lower bound

$$a_{\mathrm{R}} \geq \frac{\ln\left(\frac{1}{2\ln 2}\left(1 + \frac{1-C}{\varepsilon C}\right)\right)}{\ln\left(\frac{1}{g_1}\right)}.$$

By dropping the 1 inside the logarithm in the numerator, we obtain that

$$a_{\mathrm{R}} > K_1' + K_2' \ln\left(\frac{1}{\varepsilon}\right) \tag{2.77}$$

where $K_1' = \frac{\ln\left(\frac{1}{2\ln 2}\frac{1-C}{C}\right)}{\ln\left(\frac{1}{g_1}\right)}$ and $K_2' = \frac{1}{\ln\left(\frac{1}{g_1}\right)}$. Finally, since for a full-rank parity-check matrix, the parity-check density and average right degree are related by the equality $\Delta = \left(\frac{1-R}{R}\right) a_{\mathrm{R}}$, then we obtain the following lower bound on the asymptotic parity-check density:

$$\begin{aligned} \liminf_{m\to\infty} \Delta_m &> \frac{1-(1-\varepsilon)C}{(1-\varepsilon)C}\left(K_1' + K_2' \ln\left(\frac{1}{\varepsilon}\right)\right) \\ &> \frac{K_1 + K_2 \ln\left(\frac{1}{\varepsilon}\right)}{1-\varepsilon} \end{aligned} \tag{2.78}$$

where $K_{1,2} \triangleq \frac{1-C}{C} K_{1,2}'$. For the BEC, this lower bound can be improved by using the first bound in (2.75). In this case, $g_1 = C = 1 - p$ where p designates the erasure probability of the BEC, so the additive coefficient K_1 in the RHS of (2.70) is improved to

$$K_1 = \frac{p}{1-p} \frac{\ln\left(\frac{p}{1-p}\right)}{\ln\left(\frac{1}{1-p}\right)}.$$

This concludes the proof of this theorem. ■

Remark 2.3 For a BEC, the lower bound on the asymptotic parity-check density stated in Theorem 2.5 coincides with the bound for the BEC in [81, Eq. (3)]. This lower bound was demonstrated in [81, Theorem 2.3] to be tight. This is proved by showing that the sequence of right-regular LDPC ensembles of Shokrollahi [93] is optimal in the sense that it achieves (up to a small additive coefficient) the lower bound on the asymptotic parity-check density for the BEC.

For a general MBIOS channel (other than the BEC), we show in the proof above that the preferable logarithmic growth rate of the lower bound on the parity-check density is achieved by using the bound which follows from (2.76). However, we note that the lower bound on the parity-check density which follows from (2.75) is *universal* w.r.t. all MBIOS channels with the same capacity.

Remark 2.4 The lower bound on the parity-check density in Theorem 2.5 is uniformly tighter than the one in [81, Theorem 2.1] (except for the BSC and BEC where they coincide). For a proof of this claim, the reader is referred to Appendix 2.C.1.

Based on the proof of Theorem 2.5, we prove and discuss an upper bound on the asymptotic rate of every sequence of binary linear codes for which reliable communication is achievable. The bound refers to ML decoding, and is therefore valid for any sub-optimal decoding algorithm. Hence the following result also provides an upper bound on the achievable rate of ensembles of LDPC codes under iterative decoding, where the transmission takes places over an MBIOS channel.

Corollary 2.3 [Upper Bound on Achievable Rates] Let $\{C_m\}$ be a sequence of binary linear block codes whose codewords are transmitted with equal probability over an MBIOS channel, and assume that the block lengths of these codes tend to infinity as $m \to \infty$. Let $\Gamma_{k,m}$ be the fraction of the parity-check nodes of degree k for arbitrary representations of the codes $\mathcal{C}_m$ by bipartite graphs which corresponds to a full-rank parity-check matrix, and assume the limit $\Gamma_k \triangleq \lim_{m\to\infty} \Gamma_{k,m}$ exits. Then in the limit where $m \to \infty$, a necessary condition on the asymptotic achievable rate (R) for obtaining vanishing bit error probability is

$$R \leq 1 - \frac{1-C}{1 - \dfrac{1}{2\ln 2} \displaystyle\sum_{p=1}^{\infty} \frac{\Gamma(g_p)}{p(2p-1)}} \tag{2.79}$$

where $\Gamma(x) \triangleq \sum_k \Gamma_k x^k$, and g_p is given in (2.57).

Proof: This upper bound on the achievable rate follows immediately from Lemma 2.2 (see p. 33) and the lower bound on the conditional entropy in Proposition 2.3. The

upper bound on R follows since the bit error probability of the sequence of codes $\{\mathcal{C}_m\}$ vanishes as we let m tend to infinity. ■

Remark 2.5 We note that the upper bound on the achievable rate in the RHS of (2.79) doesn't involve maximization, in contrast to the bound in the RHS of (2.55). The second term of the maximization in the latter bound follows from considerations related to the BEC where such an expression is not required in the RHS of (2.79). The reader is referred to Appendix 2.C.2 for a proof of this claim.

Corollary 2.4 [Lower Bounds on the Bit Error Probability of LDPC Codes] Let $\mathcal{C}$ be a binary linear block code of rate R whose transmission takes place over an MBIOS channel with capacity C. For an arbitrary full-rank parity-check matrix H of the code $\mathcal{C}$, let Γ_k designate the fraction of parity-check equations that involve k variables, and $\Gamma(x) \triangleq \sum_k \Gamma_k x^k$ be the right degree distribution from the node perspective which refers to the corresponding bipartite graph of $\mathcal{C}$. Then, under ML decoding (or any other decoding algorithm), the bit error probability (P_{b}) of the code satisfies

$$h_2(P_{\mathrm{b}}) \geq 1 - \frac{C}{R} + \frac{1-R}{2R\ln 2} \sum_{p=1}^{\infty} \frac{\Gamma(g_p)}{p(2p-1)} \tag{2.80}$$

where g_p is introduced in (2.57).

Proof: This follows directly by combining (2.25) and (2.56). ■

We now introduce the definition of normalized parity-check density from [81], and derive an improved lower bound on the bit error probability (as compared to [81, Theorem 2.5]) in terms of this quantity.

Definition 2.4 [Normalized parity-check density [81]] Let $\mathcal{C}$ be a binary linear code of rate R, which is represented by a parity-check matrix H whose density is Δ. The *normalized density* of H, call it $t = t(H)$, is defined to be $t = \frac{R\Delta}{2-R}$.

In the following, we clarify the motivation for the definition of a normalized parity-check density. Let us assume that $\mathcal{C}$ is a binary linear block code of length n and rate R, and suppose that it can be represented by a bipartite graph which is *cycle-free*. From [81, Lemma 2.1], since this bipartite graph contains $(2-R)n - 1$ edges, connecting n variable nodes with $(1-R)n$ parity-check nodes without any cycles, then the parity-check density of such a cycle-free code is $\Delta = \frac{2-R}{R} - \frac{1}{nR}$. Hence, in the limit where we let n tend to infinity, the normalized parity-check density of a cycle-free code tends to 1. For codes which are represented by bipartite graphs with

cycles, the normalized parity-check density is above 1. As shown in [81, Corollary 2.5], the number of fundamental cycles in a bipartite graph which represents an arbitrary linear block $\mathcal{C}$ grows linearly with the normalized parity-check density. The normalized parity-check density therefore provides a measure for the number of cycles in bipartite graphs representing linear block codes. It is well known that cycle-free codes are not good in terms of performance, even under ML decoding [103]; hence, good error-correcting codes (e.g., LDPC codes) should be represented by bipartite graphs with cycles. Following the lead of [81], providing a lower bound on the asymptotic normalized parity-check density in terms of their rate and gap to capacity gives a quantitative measure for the number of fundamental cycles of bipartite graphs representing good error correcting codes. In the following, we provide such an improved bound as compared to the bound given in [81, Theorem 2.5]. In the continuation (see Section 2.5.2), the resulting improvement is exemplified.

First, we note that from Definition 2.4, it follows that the relation between the normalized density of a full-rank parity-check matrix and the corresponding average right degree is $t = \left(\frac{1-R}{2-R}\right) a_{\mathrm{R}}$ so the normalized parity-check density grows linearly with the average right degree (which is directly linked to the decoding complexity per iteration of LDPC codes under message-passing decoding) where the scaling factor depends on the code rate R.

Since $\sum_k k\Gamma_k = a_{\mathrm{R}}$, then by applying Jensen's inequality to the RHS of (2.80), we get the following lower bound on the bit error probability:

$$h_2(P_{\mathrm{b}}) \geq 1 - \frac{C}{R} + \frac{1-R}{2R\ln 2} \sum_{p=1}^{\infty} \frac{g_p^{\frac{(2-R)t}{1-R}}}{p(2p-1)}. \tag{2.81}$$

This lower bound on the bit error probability is tighter than the bound given in [81, Eq. (23)] because of two reasons: Firstly, by combining inequality (2.76) with the inequality proved in Appendix 2.C.1, we obtain that

$$\frac{1}{2\ln 2} \sum_{p=1}^{\infty} \frac{g_p^{\frac{(2-R)t}{1-R}}}{p(2p-1)} \geq \frac{(1-2w)^{\frac{2(2-R)t}{1-R}}}{2\ln 2}.$$

Secondly, the further improvement in the tightness of the new bound is obtained by dividing the RHS of (2.81) by R (where $R \leq 1$), as compared to the RHS of [81, Eq. (23)].

The bounds in (2.80) and (2.81) become trivial when the RHS of these inequalities are non-positive. Let the (multiplicative) gap to capacity be defined as $\varepsilon \triangleq 1 - \frac{R}{C}$. Analysis shows that the bounds in (2.80) and (2.81) are useful unless $\varepsilon \geq \varepsilon_0$. For the

bound in the RHS of (2.80), ε_0 gets the form

$$\varepsilon_0 = \frac{(1-C)B}{C(1-B)}, \quad B \triangleq \frac{1}{2\ln 2}\sum_{p=1}^{\infty}\frac{\Gamma(g_p)}{p(2p-1)} \tag{2.82}$$

and for the bound in the RHS of (2.81), ε_0 is the unique solution of the equation

$$-\varepsilon_0 C + \frac{1-(1-\varepsilon_0)C}{2\ln 2}\sum_{p=1}^{\infty}\frac{g_p^{\frac{(2-(1-\varepsilon_0)C)t}{1-(1-\varepsilon_0)C}}}{p(2p-1)} = 0. \tag{2.83}$$

For a proof of (2.82) and (2.83), the reader is referred to Appendices 2.C.3 and 2.C.4, respectively. Similarly to [81, Eq. (25)], we note that ε_0 in (2.83) forms a lower bound on the gap to capacity for an arbitrary sequence of binary linear block codes achieving vanishing bit error probability over an MBIOS channel; the bound is expressed in terms of their asymptotic rate R and normalized parity-check density t. It follows from the transition from (2.80) to (2.81) that the lower bound on the gap to capacity in (2.83) is looser as compared to the one given in (2.82). However, the bound in (2.83) solely depends on the normalized parity-check density, while the bound in (2.82) requires full knowledge of the degree distribution for the parity-check nodes.

Discussion: Due to the symmetry property, one can show that for any MBIOS channel, the conditional probability density function of the random variable $T = \Pr(X=1|Y) - \Pr(X=-1|Y)$ satisfies the property

$$f_T(t) = f_T(-t)\left(\frac{1+t}{1-t}\right), \quad -1 \le t \le 1.$$

The relation $T = \tanh\left(\frac{L}{2}\right)$ forms a one-to-one correspondence between the value of the random variable T and the value of the random variable L for the LLR. This implies that $I(X;L) = I(X;T)$. It was shown in [91, Lemma 2.4] that the moments of T satisfy the property

$$\mathbb{E}[T^{2p}] = \mathbb{E}[T^{2p-1}], \quad p \in \mathbb{N}.$$

Based on the above equality, it follows (see [91, Proposition 2.6]) that the mutual information between X and L can be expressed in the following form:

$$\begin{aligned} I(X;L) &= I(X;T) \\ &= \int_{-1}^{1} f_T(t)\log_2(1+t)\mathrm{d}t \\ &= \frac{1}{\ln 2}\sum_{p=1}^{\infty}\frac{1}{2p(2p-1)}\,\mathbb{E}[T^{2p}] \\ &= \frac{1}{\ln 2}\sum_{p=1}^{\infty}\frac{g_p}{2p(2p-1)} \end{aligned}$$

where the last transition follows from (2.57) and the symmetry property, giving the equality

$$g_p = \int_{-\infty}^{\infty} a(l) \tanh^{2p}\left(\frac{l}{2}\right) \mathrm{d}l = \mathbb{E}[T^{2p}].$$

Therefore, the mutual information is expressed as a sum of even moments of T (exactly like the lower bound on the conditional entropy in (2.56)). Similar properties also appear in [62, Lemma 3]. This gives insight to the reason for being able to express the bound on the entropy in (2.56) as a sum of one-dimensional integrals with *even* moments of the tangent hyperbolic function.

2.5 Numerical Results

We present here numerical results for the information-theoretic bounds derived in Sections 2.3 and 2.4. As expected, they significantly improve the numerical results presented in [17, Section 4] and [81, Section 4]. This improvement is attributed to the fact that, in contrast to [17, 81], in the derivation of the bounds in this paper, we do not perform a two-level quantization of the LLR which in essence converts the arbitrary MBIOS channel (whose output may be continuous) to a BSC. Throughout this section, we assume transmission of the codes over the binary-input AWGN channel.

We note that the statements in Sections 2.2–2.4 refer to the case where the parity-check matrices are full rank. Though it seems like a feasible requirement for specific linear codes, this poses a problem when considering ensembles of LDPC codes. In the latter case, a parity-check matrix, referring to a randomly chosen bipartite graph with a given pair of degree distributions, may not be full rank (one *can* even construct LDPC ensembles where the design rate is strictly less than their asymptotic rate as the block length goes to infinity). Considering ensembles of LDPC codes, it follows from the proof of Propositions 2.1, 2.2 and 2.3 that the statements stay valid with the following modifications: the actual code rate R of a code which is randomly picked from the ensemble is replaced by the design rate (R_{d}) of the ensemble, and $\{\Gamma_k\}$ becomes the degree distribution of the parity-check nodes referring to the original bipartite graph which represents a parity-check matrix, possibly not of full rank. The reason for the validity of the bound with the suggested modification is that instead of bounding the entropy of the syndrome by the sum of the entropies of its $n(1-R)$ independent components, we sum over *all* $(1-R_{\mathrm{d}})n$ components (where since $R_{\mathrm{d}} \leq R$, some of these components are possibly linearly dependent). It follows from the proofs that the entropy of the transmitted codeword ($\mathbf{X}$) cancels out with the entropy of the index L of the received vector in the appropriate coset (see e.g. (2.19)), regardless

of the rank of H. By doing this modification, the bound becomes looser when the asymptotic rate of the codes is strictly above the design rate of the ensemble. In light of this modification, we note that the fraction Γ_k of nodes of degree k is calculated in terms of the degree distribution ρ by equation (2.6), which gives

$$\Gamma_k = \frac{\rho_k}{k} \frac{1}{\int_0^1 \rho(x)\,\mathrm{d}x}\,.$$

Though the bounds in Sections 2.3–2.4 improve on the bounds in [17], this conclusion about the possible replacement of the code rate with the design rate in case that the parity-check matrices are not full rank was also noted in [17, p. 2439].

Based on [60, Lemma 7] (see Lemma 2.1 on p. 25), it was verified that the design rates of the LDPC ensembles presented in this section are equal with probability 1 to the asymptotic rates of codes from these ensembles. This allows one to consider the Shannon capacity limit for these ensembles by referring to the capacity values which correspond to their design rates (see Tables 2.1–2.3).

2.5.1 Thresholds of LDPC Ensembles under ML Decoding

Tables 2.1–2.3 provide bounds on the thresholds of LDPC ensembles under ML decoding. They also give an indication on the inherent loss in performance due to the sub-optimality of iterative decoding.

LDPC Ens.	Capacity Limit	Lower Bounds				Upper Bound [41]	DE Threshold
		2-Level	4-Level	8-Level	Un-Quantized		
(3,6)	+0.187 dB	+0.249 dB	+0.332 dB	+0.361 dB	+0.371 dB	+0.673 dB	+1.110 dB
(4,6)	−0.495 dB	−0.488 dB	−0.472 dB	−0.463 dB	−0.463 dB	−0.423 dB	+1.674 dB
(3,4)	−0.794 dB	−0.761 dB	−0.713 dB	−0.694 dB	−0.687 dB	−0.510 dB	+1.003 dB

Table 2.1: Comparison of thresholds for Gallager's ensembles of regular LDPC codes transmitted over the binary-input AWGN channel. The 2-level lower bound on the threshold of $\frac{E_\mathrm{b}}{N_\mathrm{o}}$ refers to ML decoding, and is based on [17, Theorem 1] (see also [81, Table II]). The 4-level, 8-level and un-quantized lower bounds apply to ML decoding, and are based on Corollaries 2.1, 2.2 and 2.3, respectively. The upper bound on the threshold of $\frac{E_\mathrm{b}}{N_\mathrm{o}}$ holds under 'typical pairs' decoding [41] (and hence, also under ML decoding). The DE thresholds are based on the density evolution analysis, providing exact thresholds under the iterative sum-product decoding algorithm [73].

The bounds on the achievable rates derived in [17] and Corollaries 2.1, 2.2 and 2.3 provide lower bounds on the $\frac{E_\mathrm{b}}{N_0}$ thresholds under ML decoding. For Gallager's regular

LDPC ensembles, the gap between the thresholds under ML decoding and the exact thresholds under the sum-product decoding algorithm (which are calculated using density-evolution analysis) are rather large. For this reason, we also compare the lower bounds on the $\frac{E_b}{N_0}$ thresholds under ML decoding with upper bounds on the $\frac{E_b}{N_0}$ thresholds which rely on "typical pairs decoding" [41]; an upper bound on the $\frac{E_b}{N_0}$ thresholds under an arbitrary sub-optimal decoding algorithm (e.g., "typical pairs decoding") also forms an upper bound on these thresholds under ML decoding. It is shown in Table 2.1 that for regular LDPC ensembles, the gap between the thresholds under iterative sum-product decoding and ML decoding is rather large (this follows by comparing the columns referring to the DE threshold and the upper bound based on "typical pairs decoding"). This large gap is attributed to the sub-optimality of belief propagation decoding for regular LDPC ensembles. On the other hand, it is also demonstrated in Table 2.1 that the gap between the upper and lower bounds on the thresholds under ML decoding is much smaller. For example, according to the numerical results in Table 2.1, the inherent loss in the asymptotic performance due to the sub-optimality of belief propagation for Gallager's ensemble of $(4, 6)$ regular LDPC codes (whose design rate is $\frac{1}{3}$ bits per channel use) ranges between 2.097 and 2.137 dB.

$\lambda(x)$	$\rho(x)$	Lower Bounds				DE Threshold
		2-Level	4-Level	8-Level	Un-Quantized	
$0.38354x + 0.04237x^2 + 0.57409x^3$	$0.24123x^4 + 0.75877x^5$	0.269 dB	0.370 dB	0.404 dB	0.417 dB	0.809 dB
$0.23802x + 0.20997x^2 + 0.03492x^3 + 0.12015x^4 + 0.01587x^6 + 0.00480x^{13} + 0.37627x^{14}$	$0.98013x^7 + 0.01987x^8$	0.201 dB	0.226 dB	0.236 dB	0.239 dB	0.335 dB
$0.21991x + 0.23328x^2 + 0.02058x^3 + 0.08543x^5 + 0.06540x^6 + 0.04767x^7 + 0.01912x^8 + 0.08064x^{18} + 0.22798x^{19}$	$0.64854x^7 + 0.34747x^8 + 0.00399x^9$	0.198 dB	0.221 dB	0.229 dB	0.232 dB	0.310 dB
$0.19606x + 0.24039x^2 + 0.00228x^5 + 0.05516x^6 + 0.16602x^7 + 0.04088x^8 + 0.01064x^9 + 0.00221x^{27} + 0.28636x^{29}$	$0.00749x^7 + 0.99101x^8 + 0.00150x^9$	0.194 dB	0.208 dB	0.214 dB	0.216 dB	0.274 dB

Table 2.2: Comparison of thresholds for rate one-half ensembles of irregular LDPC codes transmitted over the binary-input AWGN channel. The Shannon capacity limit corresponds to $\frac{E_b}{N_o} = 0.187$ dB. The 2-level, 4-level, 8-level and un-quantized lower bounds on the threshold refer to ML decoding, and are based on [17, Theorem 2], Corollaries 2.1, 2.2 and 2.3, respectively. The degree distributions of the ensembles and their DE thresholds are based on density evolution analysis under iterative sum-product decoding [73], and are taken from [71, Tables 1 and 2].

$\lambda(x)$	$\rho(x)$	Lower Bounds				DE
		2-Level	4-Level	8-Level	Un-Quantized	Threshold
$0.302468x + 0.319447x^2 + 0.378085x^4$	x^{11}	1.698 dB	1.786 dB	1.815 dB	1.825 dB	2.049 dB
$0.244067x + 0.292375x^2 + 0.463558x^6$	x^{13}	1.664 dB	1.718 dB	1.736 dB	1.742 dB	1.874 dB
$0.205439x + 0.255432x^2 + 0.0751187x^4 + 0.1013440x^5 + 0.3626670x^{11}$	x^{15}	1.647 dB	1.680 dB	1.691 dB	1.695 dB	1.763 dB

Table 2.3: Comparison of thresholds for rate-$\frac{3}{4}$ ensembles of irregular LDPC codes transmitted over the binary-input AWGN channel. The Shannon capacity limit corresponds to $\frac{E_b}{N_o} = 1.626$ dB. The 2-level, 4-level, 8-level and un-quantized lower bounds on the threshold refer to ML decoding, and are based on [17, Theorem 2], Corollaries 2.1, 2.2 and 2.3, respectively. The degree distributions of the ensembles and their DE thresholds are based on density evolution analysis under iterative sum-product decoding [73], and are taken from [107].

For carefully chosen ensembles of LDPC codes, it is shown in Tables 2.2 and 2.3 that the gap between the DE thresholds under the sum-product decoding algorithm and the improved lower bounds on the $\frac{E_b}{N_0}$ thresholds derived in this paper is already rather small. This indicates that for the degree distributions which are provided by the LDPC optimizer [107], the asymptotic degradation in performance due to the sub-optimality of belief propagation is marginal (it is observed from Tables 2.2 and 2.3 that for several LDPC ensembles, this degradation in the asymptotic performance is at most in the order of hundredths of a decibel).

The plots in Figure 2.2 compare different lower bounds on the $\frac{E_b}{N_0}$-threshold under ML decoding of right-regular LDPC ensembles. The plots refer to a right degree of 6 (upper plot) or 10 (lower plot). The following lower bounds are depicted in these plots: the Shannon capacity limit, the 2-level quantization lower bound in [17, Theorem 1], the 4 and 8-level quantization bounds of the LLR in Section 2.3, and finally, the bound in Section 2.4 where no quantization of the LLR is performed. It can be observed from the two plots in Figure 2.2 that the range of code rates where there exists a visible improvement with the new lower bounds depends on the degree of the parity-check nodes. In principle, the larger the value of the right-degree is, then the improvement obtained by these bounds is more pronounced starting from a higher rate code rate (e.g., for a right degree of 6 or 10, the improvement obtained by the new bounds is observed for code rates starting from 0.35 and 0.55 bits per channel use, respectively).

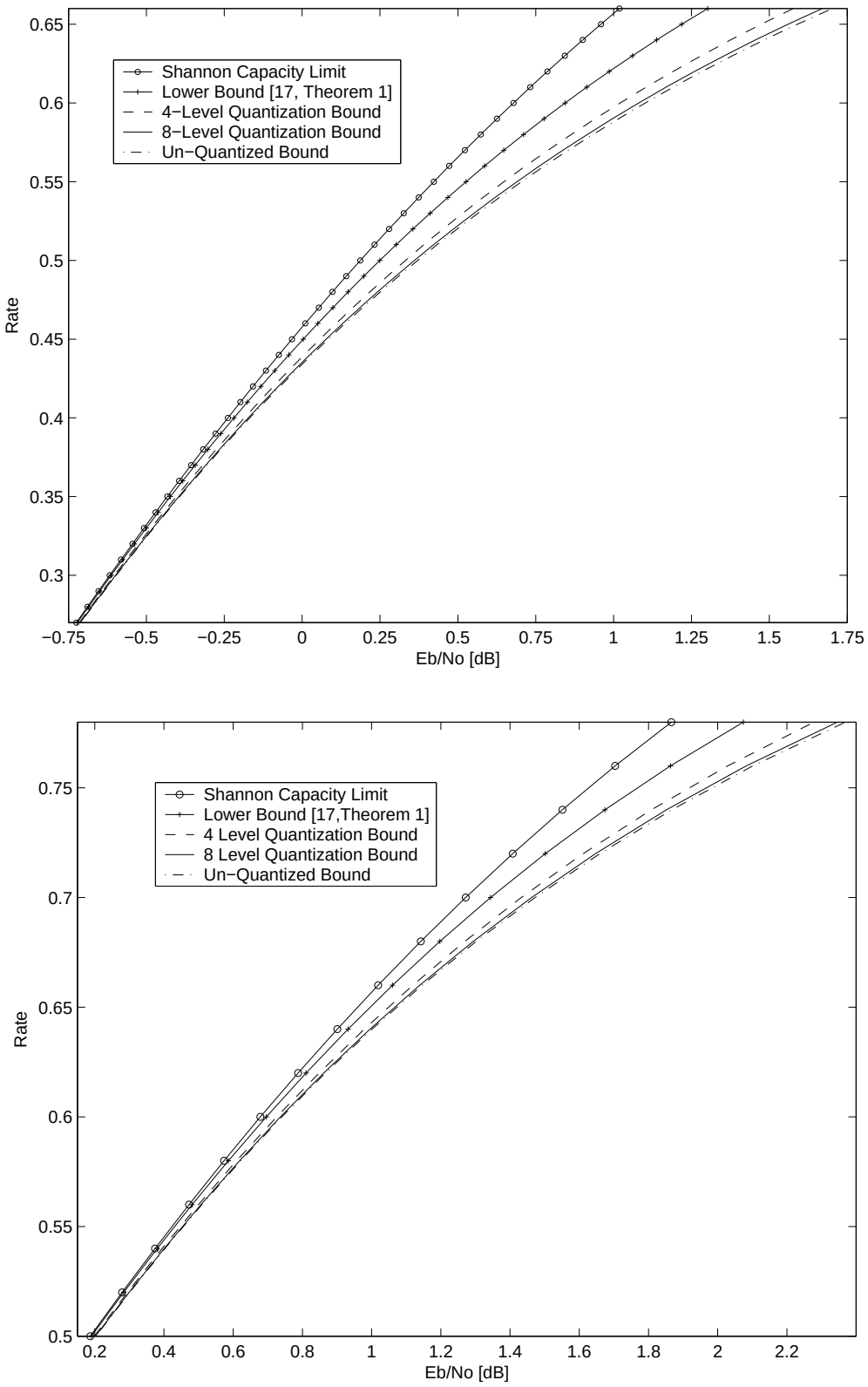

Figure 2.2: Comparison between lower bounds on the $\frac{E_\mathrm{b}}{N_0}$–thresholds under ML decoding for right-regular LDPC ensembles with $a_\mathrm{R} = 6$ (upper plot) and $a_\mathrm{R} = 10$ (lower plot). The transmission takes place over the binary-input AWGN channel.

2.5.2 Lower Bounds on the Bit Error Probability of LDPC Codes

By combining the lower bound in Proposition 2.3 and Lemma 2.2, we obtain in Corollary 2.4 an improved lower bound on the bit error probability of binary linear block codes, as compared to the one given in [81, Theorem 2.5]. The plot of Figure 2.3 presents a comparison of these lower bounds for binary linear block codes where the bounds rely on (2.81) and [81, Theorem 2.5]. They are plotted as a function of the

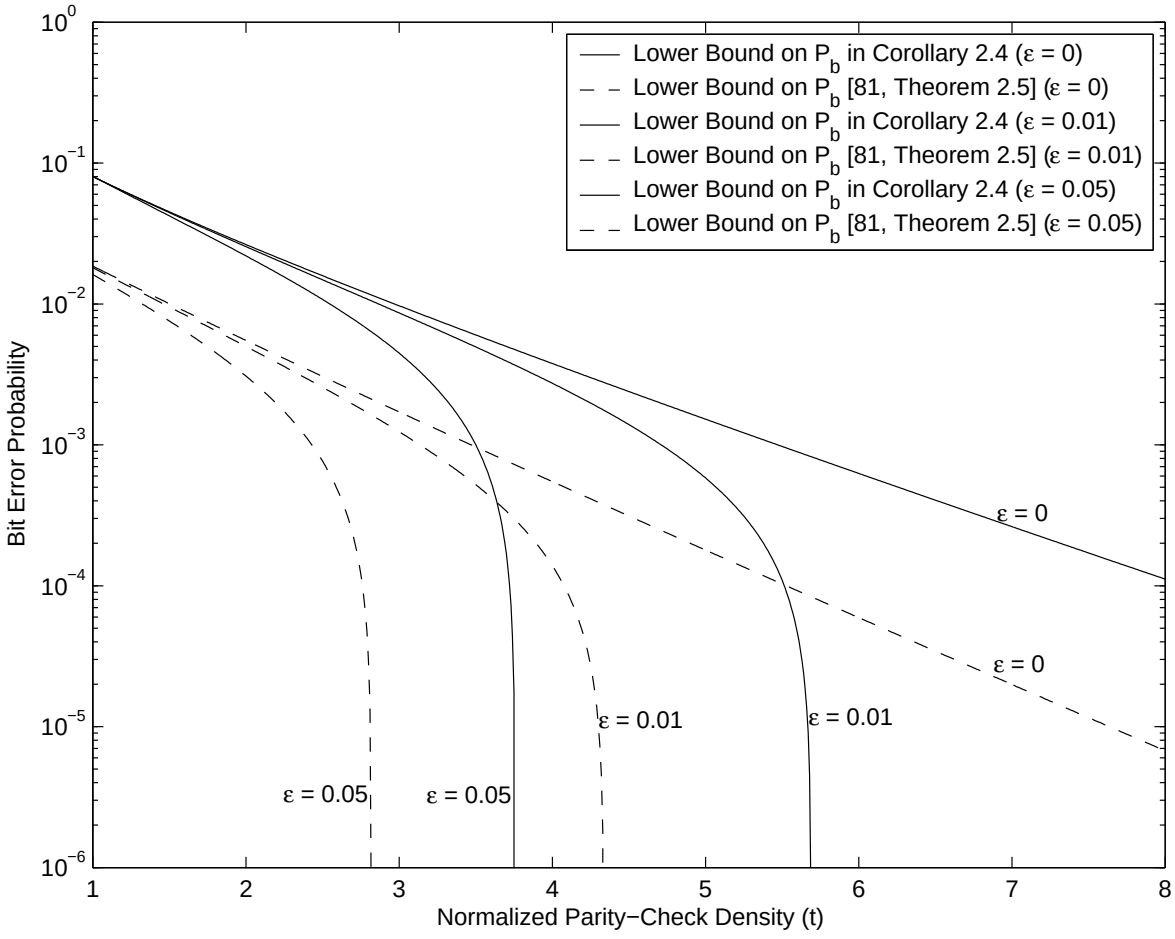

Figure 2.3: Lower bounds on the bit error probability for any binary linear block code transmitted over a binary-input AWGN channel whose capacity is $\frac{1}{2}$ bits per channel use. The bounds are depicted in terms of the normalized density of an arbitrary parity-check matrix which represents the code, and the curves correspond to code rates which are a fraction $1 - \varepsilon$ of the channel capacity (for different values of ε). The bounds depicted in dashed lines are based on [81, Theorem 2.5], and the bounds in solid lines are given in Corollary 2.4.

normalized density of an arbitrary parity-check matrix (see definition 2.4). In our setting, the capacity of the channel is $\frac{1}{2}$ bit per channel use, and the bounds are depicted for binary linear block codes whose rate is a fraction $1-\varepsilon$ of the channel capacity. To demonstrate the advantage of the lower bound on the bit error probability in (2.81) over the lower bound derived in [81, Theorem 2.5], let us assume that one wants to design a binary LDPC code which achieves a bit-error probability of 10^{-6} at a rate which is 99% of the channel capacity. The curve of the lower bound from [81] for $\varepsilon = 0.01$ implies that the normalized density of an arbitrary parity-check matrix which represents the code (see Definition 2.4 on p. 55) should be at least 4.33, while the curve depicting the bound from (2.81) strengthens this requirement to a normalized density (of each parity-check matrix) of at least 5.68. Translating this into terms of parity-check density (which is also the complexity per iteration for message-passing decoding) yields minimal parity-check densities of 13.16 and 17.27, respectively (the minimal parity-check density is given by $\Delta_{\min} = \frac{(2-R)t_{\min}}{R}$). It is reflected from Figure 2.3 that as the gap to capacity ε tends to zero, the lower bound on the normalized density of an arbitrary parity-check matrix (t), which represents a code which achieves low error probability for a rate of $R = (1-\varepsilon)C$ grows significantly.

2.5.3 Lower Bounds on the Asymptotic Parity-Check Density

The lower bound on the parity-check density derived in Theorem 2.5 enables to assess the tradeoff between asymptotic performance and asymptotic decoding complexity (per iteration) of an iterative message-passing decoder. This bound tightens the lower bound on the asymptotic parity-check density derived in [81, Theorem 2.1]. Figure 2.4 compares these bounds for codes of rate $\frac{1}{2}$ (left plot) and $\frac{3}{4}$ (right plot) where the bounds are plotted as a function of $\frac{E_{\mathrm{b}}}{N_0}$. It can be observed from Figure 2.4 that as $\frac{E_{\mathrm{b}}}{N_0}$ increases, the advantage of the bound in Theorem 2.5 over the bound in [81, Theorem 2.1] diminishes. This follows from the fact that as the value of $\frac{E_{\mathrm{b}}}{N_0}$ is increased, the two-level quantization of the LLR used in [17] and [81, Theorem 2.1] better captures the true behavior of the MBIOS channel. It is also reflected in this figure that as ε tends to zero (i.e., when the gap to capacity vanishes), the slope of the bounds becomes very sharp. This is due to the logarithmic behavior of the bounds.

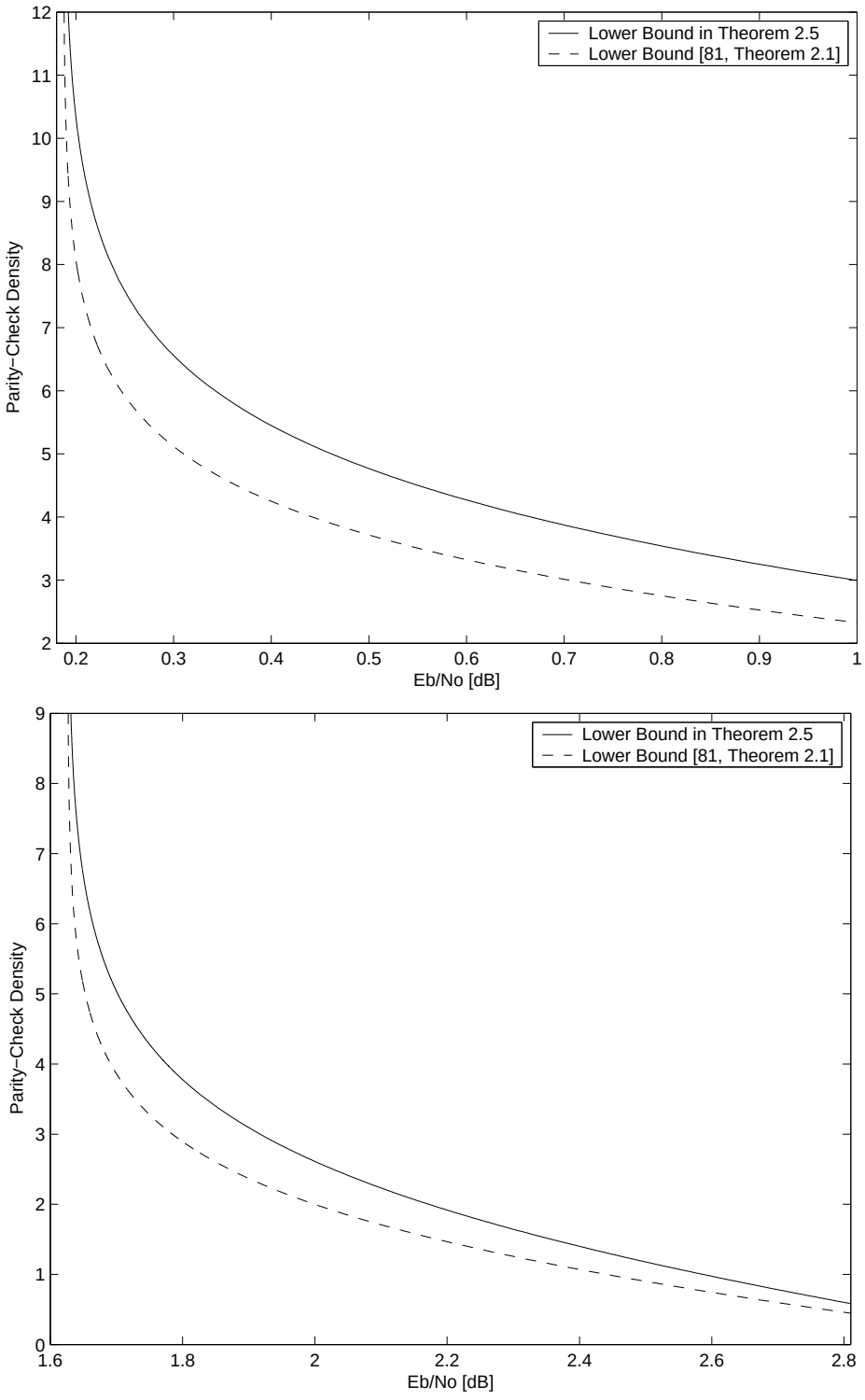

Figure 2.4: Comparison between lower bounds on the asymptotic parity-check density of binary linear block codes where the transmission takes place over a binary-input AWGN channel. The dashed line refers to [81, Theorem 2.1], and the solid line refers to Theorem 2.5. The left and right plots refer to code rates of $\frac{1}{2}$ and $\frac{3}{4}$, respectively. The Shannon capacity limit for these code rates corresponds to $\frac{E_{\mathrm{b}}}{N_0}$ of 0.187 dB and 1.626 dB, respectively.

2.6 Summary and Outlook

The outstanding performance of low-density parity-check (LDPC) codes under iterative decoding is attributed to the sparseness of the parity-check matrices of these codes. Motivated to consider how sparse parity-check matrices of binary linear block codes can be as a function of their achievable rates and their gap to capacity, we derive in this paper two kinds of bounds. The first category is some improved lower bounds on the asymptotic density of parity-check matrices in terms of the gap to capacity, and the second category is upper bounds on the achievable rates of binary linear block codes (even under ML decoding). These bounds refer to the case where the transmission takes place over memoryless binary-input output-symmetric (MBIOS) channels, and improve the tightness of the bounds given in [17, 81] (as exemplified in Section 2.5).

The information-theoretic bounds are valid for *every* sequence of binary linear block codes, in contrast to high probability results which follow from probabilistic tools (e.g., density evolution (DE) analysis under iterative message-passing decoding). The bounds hold under maximum-likelihood (ML) decoding, and hence, they hold in particular under any sub-optimal decoding algorithm. We apply the bounds to ensembles of LDPC codes where the significance of these bounds is as follows: Firstly, by comparing the new upper bounds on the achievable rates with thresholds provided by DE analysis, we obtain rigorous bounds on the loss in performance of various LDPC ensembles due to the sub-optimality of message-passing decoding (as compared to ML decoding). Secondly, the parity-check density of binary linear block codes which are represented by standard bipartite graphs can be interpreted as the complexity per iteration under message-passing decoding. Therefore, by tightening the reported lower bound on the asymptotic parity-check density (see [81, Theorem 2.1]), the new bounds provide better insight on the tradeoff between the asymptotic performance and the asymptotic decoding complexity of iteratively decoded LDPC codes. Thirdly, the new lower bound on the bit error probability of binary linear block codes tightens the reported lower bound in [81, Theorem 2.5] and provides a quantitative measure to the number of fundamental cycles in the graph which should exist in terms of the achievable rate (even under ML decoding) and its gap to capacity. It is well known that cycle-free codes have poor performance [103], so the lower bound on the minimal number of fundamental cycles in the graph (i.e., cycles which cannot be decomposed into some more elementary cycles) as a function of the gap in rate to capacity strengthens the result in [103] on the inherent limitation of cycle-free codes.

The derivation of the bounds in Section 2.3 was motivated by the desire to generalize the results in [17, Theorems 1 and 2] and [81, Theorem 2.1]. The two-level quantization of the log-likelihood ratio (LLR) which in essence replaces the arbitrary MBIOS channel by a physically degraded binary symmetric channel (BSC), is modified in Section 2.3 to a quantized channel which better reflects the statistics of the original channel (though the quantized channel is still physically degraded w.r.t. the original channel). The number of quantization levels at the output of the new channel is an arbitrary integer power of 2. The calculation of the bounds in Section 2.3 is subject to an optimization of the quantization levels of the LLR, as to get the tightest bounds within their form. In Section 2.4, we rely on the conditional pdf of the LLR at the output of the MBIOS channel, and operate on an equivalent channel without quantizing the LLR. This second approach finally leads to bounds which are uniformly tighter than the bounds in Section 2.3. It appears to be even simpler to calculate the un-quantized bounds in Section 2.4, as their calculation does not involve the solution of any optimization equation related to the quantization levels. The comparison between the quantized and un-quantized bounds gives insight on the effect of the number of quantization levels of the LLR (even if they are chosen optimally) on the achievable rates, as compared to the ideal case where no quantization is done. The results of such a comparison are shown in Tables 2.1–2.3 (see Section 2.5.1), and indicate that the improvement in the tightness of the bounds when more than 8 levels of quantization are used is marginal (in the case that the quantization levels are optimally determined). We also note that practically, the possibility to calculate un-quantized bounds which are uniformly better than the quantized bounds was facilitated due to an efficient transformation of the multi-dimensional integral in Appendix 2.B.2 into an infinite series of one-dimensional integrals whose convergence rate is fast.

In [62], a new method for analyzing LDPC and low-density generator-matrix (LDGM) codes under MAP decoding is introduced, based on tools from statistical physics. It allows to construct lower bounds on the entropy of the transmitted message given the received one. This bound involves the calculation of a supremum over all probability densities of an expression, given as a sum of statistical expectations with a number of terms growing exponentially with the maximal right and left degrees (see [62, Eqs. (6.2) and (6.3)]); this imposes a considerable difficulty in their calculation. It is of theoretical interest to see if the lower bound on the conditional entropy given in (2.56) could be re-derived by plugging a specific density in [62, Eq. (6.2)] (this question was posed in [61]). However, it is noted that the bound in (2.56) is simple to calculate. Naturally, the optimized density in the bound in [62, Eqs. (6.2) and (6.3)]

depends on the pair of degree distributions and the communication channel. Even if it could be theoretically derived as a particular case of the general bound given in [62], the simplicity of the calculation of the information-theoretic bounds in Sections 2.3 and 2.4 is of practical significance, especially in light of the fact that they provide numerical results which are reasonably close to the optimized bounds in [62]. As an example, for the binary erasure channel (BEC), the lower bounds on the thresholds of various rate-$\frac{1}{2}$ regular LDPC ensembles under ML decoding were compared with those provided in [62, Table 1], and the typical differences in the thresholds for these ensembles were minor (in the order of 10^{-4}). Note that the information-theoretic bounds derived in this paper are valid for every binary linear block code (the transition to sequences of codes is later used to get an upper bound on the achievable rates and a lower bound on the asymptotic parity-check density). Considering ensembles of codes, one gets from [62] high probability results as the block length gets large, but the performance of specific codes of finite length deviates from the average ensemble performance. Since the bounding techniques which rely on statistical physics [62] do not allow for a bound which is valid for every linear block code, it would be interesting to get some theory that unifies the information-theoretic and statistical physics approaches and provides bounds that are tight on average and valid code by code.

The bounds on the thresholds of LDPC ensembles under ML decoding depend only on the degree distribution of their parity-check nodes and their design rate. For a given parity-check degree distribution (ρ) and design rate (R), the bounds provide an indication on the inherent gap to capacity which is independent of the choice of the left degree distribution λ (as long as the pair of degree distributions (λ, ρ) yield the design rate R). Sections 2.3 and 2.4 give *universal* bounds on the gap to capacity for general LDPC ensembles over MBIOS channels, no matter what the degree of the variable nodes is (as long as the rate is fixed). These bounds can be exploited to gain insight on how good a specific design of degree distributions is in terms of the design rate and the average right degree where this is done by comparing the DE thresholds and the lower bounds on the ML thresholds, as in Tables 2.2 and 2.3 in Section 2.5. On the other hand, the bounds are not necessarily tight for LDPC ensembles with a given pair of degree distributions (λ, ρ) since the explicit influence of λ is not taken into account except through the design rate of the ensemble. As a topic for further research, it is suggested to examine the possibility of tightening the bounds for specific ensembles by explicitly taking into account the exact characterization of λ (a possible direction studied by the authors is based on the analysis of the average coset weight distributions of ensembles of binary linear block codes which play a crucial role in tightening the upper bound on the syndrome entropy [112]). The numerical results

shown in Section 2.5 indicate, however, that these bounds are useful for assessing the inherent gap to capacity of various LDPC ensembles. The gap to capacity is an inherent limitation which is attributed to the finite average right degree of these LDPC ensembles [81].

As a topic for further research, we also suggest to study a possible generalization of the bounds to non-binary linear block codes. These generalized bounds can be applied to the analysis of the ML performance of non-binary LDPC ensembles whose transmission takes place over arbitrary discrete memoryless channels with possibly different types of quantization [12].

The lower bound on the asymptotic parity-check density in [81, Theorem 2.1] and its improvements in Sections 2.3 and 2.4 grow like the log of the inverse of the gap (in rate) to capacity. The result in [81, Theorem 2.2] shows that a logarithmic growth rate of the parity-check density is achievable for Gallager's regular LDPC ensemble under ML decoding when transmission takes place over an arbitrary MBIOS channel. These results show that for any iterative decoder which is based on the representation of the codes by Tanner graphs, there exists a tradeoff between asymptotic performance and complexity which cannot be surpassed. Recently, it was shown in [65] that a better tradeoff can be achieved by allowing more complicated graphical models which involve a sufficient number of state nodes in the graph; for the particular case of the BEC, the encoding and the decoding complexity of properly designed codes on graphs remains bounded as the gap to capacity vanishes (see [65]).

In [85], the authors consider the achievable rates and decoding complexity of LDPC codes over statistically independent *parallel channels*, and generalize in a non-trivial way the un-quantized bounds introduced in Section 2.4. The bounds in [85] are applied to randomly and intentionally punctured LDPC codes, and improved puncturing theorems are derived as compared to those introduced in [65, Theorems 3 and 4].

Appendices

2.A Some mathematical details related to the proofs of the statements in Section 2.3.1

2.A.1 Proof of Lemma 2.2

We prove here Lemma 2.2. Since there is a one to one correspondence between the codewords and the set of information bits used to encode them, then $H(\mathbf{X}|\mathbf{Y}) = H(\mathbf{U}|\mathbf{Y})$ where the vector $\mathbf{u} = (u_1, \ldots, u_{nR})$ denotes the sequence of information bits used to encode the codeword $\mathbf{x}$. Let $P_{\mathrm{b}}^{(i)}$ denote the probability of decoding the bit u_i erroneously given the received sequence at the output of the MBIOS channel, then the bit error probability is given by

$$P_{\mathrm{b}} = \frac{1}{nR} \sum_{i=1}^{nR} P_{\mathrm{b}}^{(i)}. \tag{2.A.1}$$

This therefore gives

$$\begin{aligned}
\frac{H(\mathbf{X}|\mathbf{Y})}{n} &= \frac{H(\mathbf{U}|\mathbf{Y})}{n} \\
&\overset{(a)}{\leq} \frac{1}{n} \sum_{i=1}^{nR} H(U_i|\mathbf{Y}) \\
&\overset{(b)}{\leq} \frac{1}{n} \sum_{i=1}^{nR} h_2\left(P_{\mathrm{b}}^{(i)}\right) \\
&\overset{(c)}{\leq} R\, h_2\left(\frac{1}{nR} \sum_{i=1}^{nR} P_{\mathrm{b}}^{(i)}\right) \\
&\overset{(d)}{=} R\, h_2\big(P_{\mathrm{b}}\big)
\end{aligned}$$

where inequality (a) holds from the chain rule of the entropy and since conditioning reduces entropy, inequality (b) follows from Fano's inequality and since the code is binary, inequality (c) is based on Jensen's inequality and the concavity of the binary entropy function (h_2), and equality (d) follows from (2.A.1).

2.A.2 Derivation of the Optimization Equation in (2.24) and Proving the Existence of its Solution

Derivation of the optimization equation (2.24): We derive here the optimization equation (2.24) which refers to the "four-level quantization" lower bound on the parity-check density (see p. 33).

Let the function a designate the conditional pdf of the LLR at the output of the original MBIOS channel, given the zero symbol is transmitted. In the following, we express the transition probabilities of the degraded channel in Figure 2.1 (see p. 28) in terms of the pdf a and the value of l:

$$p_0 = \Pr(Z = 0 \,|\, X = 0) = \int_l^{\infty} a(u)\,\mathrm{d}u \tag{2.A.2}$$

$$p_1 = \Pr(Z = \alpha \,|\, X = 0) = \int_{0^+}^{l} a(u)\,\mathrm{d}u + \frac{1}{2}\int_{0^-}^{0^+} a(u)\,\mathrm{d}u \tag{2.A.3}$$

$$p_2 = \Pr(Z = 1+\alpha \,|\, X = 0) = \int_{-l}^{0^-} a(u)\,\mathrm{d}u + \frac{1}{2}\int_{0^-}^{0^+} a(u)\,\mathrm{d}u \tag{2.A.4}$$

$$p_3 = \Pr(Z = 1 \,|\, X = 0) = \int_{-\infty}^{-l} a(u)\,\mathrm{d}u. \tag{2.A.5}$$

We note that the integration of the function a from $u = 0^-$ to $u = 0^+$ gives a non-zero value if and only if there is a non-vanishing probability that the value of the LLR at the output of the original channel is zero (e.g., a BEC). Otherwise, the contribution of this integral to (2.A.3) and (2.A.4) vanishes. Since the channel is MBIOS, the symmetry property [73] gives

$$a(u) = e^u\, a(-u), \quad \forall\, u \in \mathbb{R}. \tag{2.A.6}$$

Based on the expressions for the coefficients K_1 and K_2 in the lower bound on the asymptotic parity-check density (2.22), then in order to find the tightest lower bound then we need to maximize

$$\frac{(p_1 - p_2)^2}{p_1 + p_2} + \frac{(p_0 - p_3)^2}{p_0 + p_3} \tag{2.A.7}$$

w.r.t. the free parameter $l \in \mathbb{R}^+$. From Eqs. (2.A.2)–(2.A.5) and the symmetry property in (2.A.6)

$$p_0 - p_3 = \int_l^{\infty} a(u)(1 - e^{-u})\,\mathrm{d}u \quad \Rightarrow \quad \frac{\partial}{\partial l}(p_0 - p_3) = -a(l)(1 - e^{-l}) \tag{2.A.8}$$

$$p_0 + p_3 = \int_l^{\infty} a(u)(1 + e^{-u})\,\mathrm{d}u \quad \Rightarrow \quad \frac{\partial}{\partial l}(p_0 + p_3) = -a(l)(1 + e^{-l}) \tag{2.A.9}$$

$$p_1 - p_2 = \int_{0^+}^{l} a(u)(1 - e^{-u})\,\mathrm{d}u \quad \Rightarrow \quad \frac{\partial}{\partial l}(p_1 - p_2) = a(l)(1 - e^{-l}) \tag{2.A.10}$$

$$p_1 + p_2 = \int_{0^+}^{l} a(u)(1 + e^{-u})\,\mathrm{d}u \quad \Rightarrow \quad \frac{\partial}{\partial l}(p_1 + p_2) = a(l)(1 + e^{-l}) \tag{2.A.11}$$

so the calculation of the partial derivative of (2.A.7) w.r.t. l gives

$$\begin{aligned}
&\frac{\partial}{\partial l}\left\{\frac{(p_1 - p_2)^2}{p_1 + p_2} + \frac{(p_0 - p_3)^2}{p_0 + p_3}\right\} \\
&= -4\, a(l)\left\{\left[\left(\frac{p_2}{p_1 + p_2}\right)^2 - \left(\frac{p_3}{p_0 + p_3}\right)^2\right] + e^{-l}\left[\left(\frac{p_1}{p_1 + p_2}\right)^2 - \left(\frac{p_0}{p_0 + p_3}\right)^2\right]\right\}.
\end{aligned}$$

Since the first derivative of a function changes its sign at a neighborhood of any local maxima or minima point, and since the pdf a is always non-negative, then the second multiplicative term above is the one which changes its sign at a neighborhood of l maximizing (2.A.7). For this value of l, the second multiplicative term vanishes, which gives the optimization equation for l in (2.24).

Proof of existence of a solution to (2.24)*:* In order to show that a solution to (2.24) always exists, we will see how the LHS and the RHS of this equation behave as $l \to 0^+$ and $l \to \infty$. From (2.A.2)–(2.A.5), it follows that in the limit where $l \to \infty$, we get

$$p_1 \to 1 - w - \Pr(\mathrm{LLR}(Y) = \infty \,|\, X = 0), \quad p_2 \to w$$

where w is introduced in (2.2), and therefore

$$\lim_{l \to \infty} \frac{p_2^2 + e^{-l} p_1^2}{(p_1 + p_2)^2} = \left(\frac{w}{1 - \Pr(\mathrm{LLR}(Y) = \infty \,|\, X = 0)} \right)^2. \tag{2.A.12}$$

Since from the symmetry property

$$p_3 = \int_l^\infty a(-u) du = \int_l^\infty e^{-u} a(u) \mathrm{d}u \leq e^{-l} \int_l^\infty a(u) \mathrm{d}u = e^{-l} p_0$$

then the fraction $\frac{p_3}{p_0}$ tends to zero as $l \to \infty$, so

$$\lim_{l \to \infty} \frac{p_3^2 + e^{-l} p_0^2}{(p_0 + p_3)^2} = \lim_{l \to \infty} \frac{\left(\frac{p_3}{p_0}\right)^2 + e^{-l}}{\left(1 + \frac{p_3}{p_0}\right)^2} = 0. \tag{2.A.13}$$

It therefore follows from (2.A.12) and (2.A.13) that for large enough values of l, the LHS of (2.24) is larger than the RHS of this equation. On the other hand, in the limit where $l \to 0^+$, we get

$$p_1, p_2 \to \frac{1}{2} \int_{0^-}^{0^+} a(u) \mathrm{d}u$$

and therefore

$$\lim_{l \to 0^+} \frac{p_2^2 + e^{-l} p_1^2}{(p_1 + p_2)^2} = \frac{1}{2}. \tag{2.A.14}$$

In the limit where $l \to 0^+$

$$p_0 \to \int_{0^+}^\infty a(u) \mathrm{d}u, \quad p_3 \to \int_{-\infty}^{0^-} a(u) \mathrm{d}u, \quad p_0 + p_3 \to \beta$$

where $\beta \triangleq 1 - \int_{0^-}^{0^+} a(u) \mathrm{d}u$. By denoting $u \triangleq \int_{0^+}^\infty a(u) \mathrm{d}u$, we get $0 \leq u \leq \beta$, and

$$\lim_{l \to 0^+} \frac{p_3^2 + e^{-l} p_0^2}{(p_0 + p_3)^2} = \frac{u^2 + (\beta - u)^2}{\beta^2} \geq \frac{1}{2}, \quad \forall\, u \in [0, \beta]. \tag{2.A.15}$$

We note that the last inequality holds in equality if and only if $u = \frac{\beta}{2}$. But if this condition holds, then this implies that

$$\int_{-\infty}^{0^-} a(u)\,du = \int_{0^+}^{\infty} a(u)\,\mathrm{d}u$$

which from the symmetry property cannot be satisfied unless $a(u) = \delta(u)$. The latter condition corresponds to a BEC with erasure probability 1 (whose capacity is equal to zero).

From (2.A.14) and (2.A.15), we obtain that for small enough (and non-negative) values of l, the LHS of (2.24) is less or equal to the RHS of this equation. Since we also obtained that for large enough l, the LHS of (2.24) is larger than the RHS of this equation, the existence of a solution to (2.24) follows from continuity considerations.

2.A.3 Proof of Inequality (2.33)

We prove here the inequality (2.33) (see p. 36) which implies that the "four-level quantization" lower bound on the parity-check density (see p. 33) is tighter than what can be interpreted as the "two levels quantization" bound in [81, Theorem 2.1]. Based on (2.2), we get

$$w = \Pr\{\mathrm{LLR}(Y) < 0 \,|\, X = 0\} + \frac{1}{2}\Pr\{\mathrm{LLR}(Y) = 0 \,|\, X = 0\}$$

so from (2.9), $w = p_2 + p_3$. By invoking Jensen's inequality, we get

$$\begin{aligned}
&\frac{(p_1 - p_2)^2}{p_1 + p_2} + \frac{(p_0 - p_3)^2}{p_0 + p_3} \\
&= (p_1 + p_2)\left(\frac{p_1 - p_2}{p_1 + p_2}\right)^2 + (p_0 + p_3)\left(\frac{p_0 - p_3}{p_0 + p_3}\right)^2 \\
&\geq \left[(p_1 + p_2)\left(\frac{p_1 - p_2}{p_1 + p_2}\right) + (p_0 + p_3)\left(\frac{p_0 - p_3}{p_0 + p_3}\right)\right]^2 \\
&= (p_0 + p_1 - p_2 - p_3)^2 \\
&= (1 - 2p_2 - 2p_3)^2 \\
&= (1 - 2w)^2.
\end{aligned}$$

An equality is achieved if and only if $\frac{p_1 - p_2}{p_1 + p_2} = \frac{p_0 - p_3}{p_0 + p_3}$. From (2.A.8)–(2.A.11), we get

$$\frac{p_1 - p_2}{p_1 + p_2} = \frac{\displaystyle\int_{0^+}^{l} a(u)(1 - e^{-u})\,\mathrm{d}u}{\displaystyle\int_{0^+}^{l} a(u)(1 + e^{-u})\,\mathrm{d}u} \leq \frac{1 - e^{-l}}{1 + e^{-l}}$$

and

$$\frac{p_0 - p_3}{p_0 + p_3} = \frac{\displaystyle\int_l^\infty a(u)(1 - e^{-u})\,\mathrm{d}u}{\displaystyle\int_l^\infty a(u)(1 + e^{-u})\,\mathrm{d}u} \geq \frac{1 - e^{-l}}{1 + e^{-l}}\,.$$

The two fractions $\frac{p_1 - p_2}{p_1 + p_2}$ and $\frac{p_0 - p_3}{p_0 + p_3}$ cannot be equal unless the LLR is either equal to l or $-l$. This makes the four-level quantization of the LLR identical to the two-level quantization used for the derivation of the original bound in [17, Theorem 2]. Equality can be also achieved if $p_1 + p_2 = 0$ or $p_0 + p_3 = 0$ which converts the channel model in Figure 2.1 (see p. 28) to a BSC.

2.B Some mathematic details for the proof of Proposition 2.3

We provide in this appendix further mathematical details related to the proof of Proposition 2.3. We note that Appendix 2.B.1 serves here as a preparatory step for the derivation in Appendix 2.B.2.

2.B.1 Power Series Expansion of the Binary Entropy Function

Lemma 2.B.1

$$h_2(x) = 1 - \frac{1}{2\ln 2} \sum_{p=1}^{\infty} \frac{(1 - 2x)^{2p}}{p(2p - 1)}\,, \quad 0 \leq x \leq 1. \tag{2.B.1}$$

Proof: We prove this by expanding the binary entropy function into a power series around $\frac{1}{2}$. The first order derivative is

$$h_2'(x) = \frac{\ln\left(\frac{1-x}{x}\right)}{\ln 2}$$

and the higher order derivatives get the form

$$h_2^{(n)}(x) = -\frac{(n-2)!}{\ln 2}\left(\frac{(-1)^n}{x^{n-1}} + \frac{1}{(1-x)^{n-1}}\right), \quad n = 2, 3, \ldots\,.$$

The derivatives of odd degree therefore vanish at $x = \frac{1}{2}$, and for an even value of $n \geq 2$

$$h_2^{(n)}\left(\frac{1}{2}\right) = -\frac{(n-2)!\,2^n}{\ln 2}.$$

This yields the following power series expansion of the binary entropy function around the point $\frac{1}{2}$:

$$\begin{aligned} h_2(x) &= 1 - \sum_{n \geq 2 \, \text{even}} \left\{ \frac{\frac{(n-2)! \, 2^n}{\ln 2}}{n!} \cdot \left(x - \frac{1}{2} \right)^n \right\} \\ &= 1 - \frac{1}{\ln 2} \sum_{n \geq 2 \, \text{even}} \frac{(2x-1)^n}{n(n-1)} \\ &= 1 - \frac{1}{2 \ln 2} \sum_{p=1}^{\infty} \frac{(2x-1)^{2p}}{p(2p-1)} \end{aligned}$$

and this power series converges for all $x \in [0, 1]$. ■

We note that since the power series in (2.B.1) has always non-negative coefficients, then its truncation always gives an upper bound on the binary entropy function, i.e.,

$$h_2(x) \leq 1 - \frac{1}{2 \ln 2} \sum_{p=1}^{m} \frac{(1-2x)^{2p}}{p(2p-1)} \qquad \forall \, x \in [0, 1], \, m \in \mathbb{N}. \tag{2.B.2}$$

The case where $m = 1$ gives the upper bound in Lemma 2.3 which is used in this paper for the derivation of the lower bounds on the parity-check density. The reason for not using a tighter version of the binary entropy function for this case was because otherwise we would get a polynomial equation for a_{R} whose solution cannot be given necessarily in closed form. As shown in Figure 2.5, the upper bound on the binary entropy function h_2 over the whole interval $[0, 1]$ is improved considerably by taking even a moderate value for m (e.g., $m = 10$ gives already a very tight upper bound on h_2 which deviates from the exact values only at a small neighborhood near the two endpoints of this interval).

2.B.2 Calculation of the Multi-Dimensional Integral in (2.65)

Based on Lemma 2.B.1 which provides a power series expansion of h_2 near the point $\frac{1}{2}$, we obtain

$$\begin{aligned} &\int_0^\infty \int_0^\infty \ldots \int_0^\infty \prod_{m=1}^{k} f_\Omega(\alpha_m) \, h_2 \left(\frac{1}{2} \left(1 - \prod_{m=1}^{k} \left(\frac{1 - e^{-\alpha_m}}{1 + e^{-\alpha_m}} \right) \right) \right) \, \mathrm{d}\alpha_1 \mathrm{d}\alpha_2 \ldots \mathrm{d}\alpha_k \\ &= 1 - \frac{1}{2 \ln 2} \sum_{p=1}^{\infty} \frac{1}{p(2p-1)} \int_0^\infty \int_0^\infty \ldots \int_0^\infty \prod_{m=1}^{k} f_\Omega(\alpha_m) \prod_{m=1}^{k} \left(\frac{1 - e^{-\alpha_m}}{1 + e^{-\alpha_m}} \right)^{2p} \mathrm{d}\alpha_1 \mathrm{d}\alpha_2 \ldots \mathrm{d}\alpha_k \\ &= 1 - \frac{1}{2 \ln 2} \sum_{p=1}^{\infty} \frac{1}{p(2p-1)} \int_0^\infty \int_0^\infty \ldots \int_0^\infty \prod_{m=1}^{k} \left(f_\Omega(\alpha_m) \, \tanh^{2p} \left(\frac{\alpha_m}{2} \right) \right) \mathrm{d}\alpha_1 \mathrm{d}\alpha_2 \ldots \mathrm{d}\alpha_k \\ &= 1 - \frac{1}{2 \ln 2} \sum_{p=1}^{\infty} \frac{1}{p(2p-1)} \left(\int_0^\infty f_\Omega(\alpha) \, \tanh^{2p} \left(\frac{\alpha}{2} \right) \, \mathrm{d}\alpha \right)^k. \end{aligned}$$

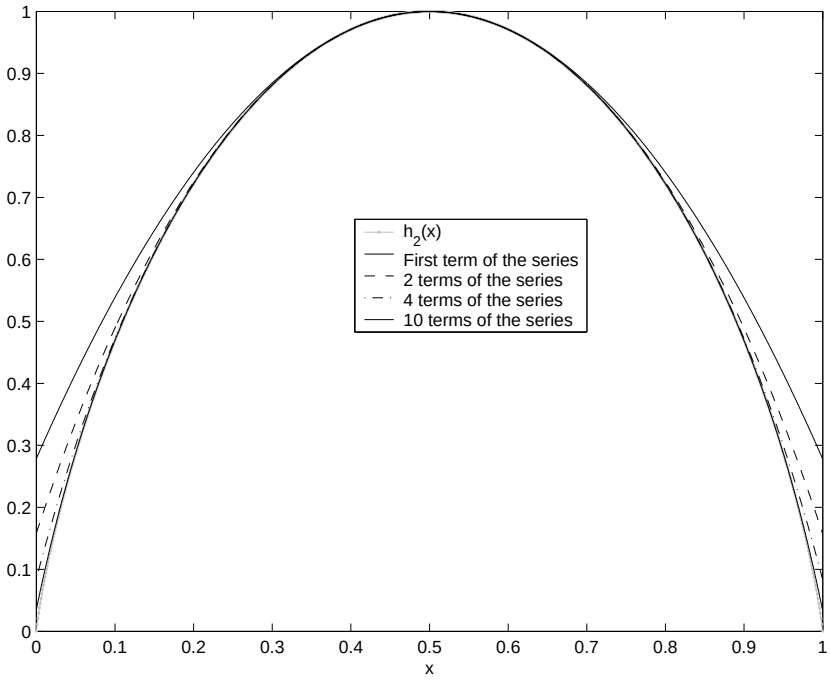

Figure 2.5: Plot of the binary entropy function to base 2 and some upper bounds which are obtained by truncating its power series around $x = \frac{1}{2}$.

This transforms the original k-dimensional integral to an infinite sum of one-dimensional integrals. Since we are interested in obtaining a tight upper bound on the k-dimensional integral above, and all the terms of the last infinite series are positive, then any truncation of the last infinite series is an upper bound. Based on the discussion in Appendix 2.B.1, we compute the first 10 terms of this series which (based on the plot in Figure 2.5) give a very tight upper bound on the k-dimensional integral (for all k).

2.C Some mathematical details related to the proofs of the statements in Section 2.4

2.C.1 On the Improved Tightness of the Lower Bound in Theorem 2.5

We show here that the lower bound on the parity-check density in Theorem 2.5 is uniformly tighter than the one in [81, Theorem 2.1] (except for the BSC and BEC where they coincide). In order to show this , we first prove the following lemma:

Lemma 2.C.1 For any MBIOS channel, $g_1 \geq (1-2w)^2$ where w and g_1 are introduced in (2.2) and (2.57), respectively.

Proof: From (2.2), (2.57) and (2.59)

$$\begin{aligned}
g_1 &= \int_0^\infty a(l)\,(1+e^{-l})\,\tanh^2\left(\frac{l}{2}\right)\,\mathrm{d}l \\
&= \int_0^\infty f_\Omega(l)\,\tanh^2\left(\frac{l}{2}\right)\,\mathrm{d}l \\
&\geq \left(\int_0^\infty f_\Omega(l)\,\tanh\left(\frac{l}{2}\right)\,\mathrm{d}l\right)^2 \\
&= \left(\int_0^\infty a(l)\,(1+e^{-l})\cdot\left(\frac{1-e^{-l}}{1+e^{-l}}\right)\,\mathrm{d}l\right)^2 \\
&= \left(\int_{0^+}^\infty a(l)\,\mathrm{d}l - \int_{0^+}^\infty e^{-l}\,a(l)\,dl\right)^2 \\
&= \left(\int_{0^+}^\infty a(l)\,\mathrm{d}l - \int_{0^+}^\infty a(-l)\,\mathrm{d}l\right)^2 \\
&= \left(\int_{0^+}^\infty [a(l)+a(-l)]\,\mathrm{d}l - 2\int_{0^+}^\infty a(-l)\,\mathrm{d}l\right)^2 \\
&= \left(\int_{-\infty}^\infty a(l)\,\mathrm{d}l - \int_{0^-}^{0^+} a(l)\,dl - 2\int_{0^+}^\infty a(-l)\,\mathrm{d}l\right)^2 \\
&= \left(1-2\left(\int_{-\infty}^{0^-} a(l)\,\mathrm{d}l + \frac{1}{2}\int_{0^-}^{0^+} a(l)\,\mathrm{d}l\right)\right)^2 \\
&= (1-2w)^2.
\end{aligned}$$

where the single inequality above follows from Jensen's inequality. ∎

The proof of the claim now follows directly by noticing that the lower bound on the parity-check density, as given in (2.70)–(2.72), is equal to

$$\frac{K_1 + K_2 \ln\left(\frac{1}{\varepsilon}\right)}{1-\varepsilon} = \frac{1-C}{(1-\varepsilon)C}\frac{\ln\left(\frac{1}{2\ln 2}\frac{1-C}{\varepsilon C}\right)}{\ln\left(\frac{1}{g_1}\right)}$$

where we refer here to all MBIOS channels except the BEC. On the other hand, the lower bound on the parity-check density which is given in [81, Theorem 2.1] gets the form

$$\frac{1-C}{(1-\varepsilon)C}\frac{\ln\left(\frac{1}{2\ln 2}\frac{1-C}{\varepsilon C}\right)}{\ln\left(\frac{1}{(1-2w)^2}\right)}\,.$$

If the lower bound is not trivial (i.e., the common numerator in both bounds is positive), then the improvement in the tightness of the former bound over the latter bound follows from Lemma 2.C.1. We note that for the particular case of a BSC, the above two bounds coincide (as for a BSC whose crossover probability is p, it is easy to verify from (2.2) and (2.57) that $w = p$ and $g_1 = (1-2p)^2$, respectively, hence $g_1 = (1-2w)^2$).

2.C.2 Proof for the Claim in Remark 2.5

In order to prove the claim in Remark 2.5 (see p. 55), it is required to show that

$$\frac{1-C}{1-\frac{1}{2\ln 2}\sum_{p=1}^{\infty}\left\{\frac{1}{p(2p-1)}\sum_{k}\Gamma_k\, g_p^k\right\}} \geq \frac{2w}{1-\sum_{k}\Gamma_k\,(1-2w)^k} \tag{2.C.1}$$

where w is introduced in (2.2). The reason for showing this in light of the claim in Remark 2.5 is that the RHS of the last inequality follows from considerations related to a BEC, essentially in the same way that the second term of the maximization in the RHS of (2.55) is derived. By showing this, we prove that the maximization of the two expressions in the LHS and RHS of (2.C.1) doesn't affect the bound in Corollary 2.3.

Following the steps which lead to (2.75), we get that for any integer $k \geq 2$

$$\frac{1}{2\ln 2}\sum_{p=1}^{\infty}\frac{g_p^k}{p(2p-1)} \geq C^k.$$

Applying this to (2.C.1) and denoting $\Gamma(x) \triangleq \sum_k \Gamma_k x^k$, we get that a sufficient condition for (2.C.1) to hold is

$$\frac{1-C}{1-\Gamma(C)} \geq \frac{2w}{1-\Gamma(1-2w)}. \tag{2.C.2}$$

From the erasure decomposition lemma, we get that an MBIOS channel is physically degraded as compared to a BEC with an erasure probability $p = 2w$. By the information processing inequality, it follows that $C \leq 1-2w$. Therefore, in order to prove (2.C.2), it is enough to show that the function

$$f(x) = \frac{1-x}{1-\Gamma(x)}$$

is monotonically decreasing for $x \in (0,1)$. We prove this property by showing that the derivative of the function f is non-positive for $x \in (0,1)$. As the denominator of the derivative is positive, we may equivalently show

$$\Gamma'(x)\,(1-x) - \big(1-\Gamma(x)\big) \leq 0.$$

Dividing both sides of the inequality by $1-x$ which is in the interval $\in (0,1)$ and noting that $\Gamma(1) = \sum_k \Gamma_k = 1$, we get that it is enough to show

$$\Gamma'(x) - \frac{\Gamma(1)-\Gamma(x)}{1-x} \leq 0. \tag{2.C.3}$$

Since the function Γ is a polynomial and therefore analytic, by the mean-value theorem we get that for some $\tilde{x} \in (x,1)$

$$\frac{\Gamma(1)-\Gamma(x)}{1-x} = \Gamma'(\tilde{x}).$$

Since $\Gamma'(x) = \sum_k k\Gamma_k x^{k-1}$ is monotonically increasing for $x \geq 0$, then (2.C.3) follows for all $x \in (0,1)$. This in turn proves (2.C.1).

2.C.3 Proof of Eq. (2.82)

In order to prove (2.82), we first multiply the two sides of (2.80) by R, and denote $R = (1-\varepsilon)C$. This gives that the lower bound on the bit error probability in (2.80) is non-positive if and only if

$$(1-C)B - \varepsilon C(1-B) \leq 0. \tag{2.C.4}$$

Unless the channel is noiseless, we get

$$\begin{aligned}
B &= \frac{1}{2\ln 2}\sum_{p=1}^{\infty}\frac{\Gamma(g_p)}{p(2p-1)} \\
&= \frac{1}{2\ln 2}\sum_{p=1}^{\infty}\left\{\frac{1}{p(2p-1)}\sum_k \Gamma_k\, g_p^k\right\} \\
&= \frac{1}{2\ln 2}\sum_{p=1}^{\infty}\left\{\frac{1}{p(2p-1)}\sum_k \Gamma_k \left(\int_{0^+}^{\infty} a(l)(1+e^{-l})\tanh^{2p}\left(\frac{l}{2}\right)\mathrm{d}l\right)^k\right\} \\
&< \frac{1}{2\ln 2}\sum_{p=1}^{\infty}\left\{\frac{1}{p(2p-1)}\sum_k \Gamma_k \left(\int_{0^+}^{\infty} a(l)(1+e^{-l})\mathrm{d}l\right)^k\right\} \\
&= \frac{1}{2\ln 2}\sum_{p=1}^{\infty}\left\{\frac{1}{p(2p-1)}\sum_k \Gamma_k \left(\int_{\mathbb{R}-\{0\}} a(l)\mathrm{d}l\right)^k\right\} \\
&= \frac{1}{2\ln 2}\sum_{p=1}^{\infty}\left\{\frac{1}{p(2p-1)}\sum_k \Gamma_k \Big(1-\Pr(\mathrm{LLR}=0)\Big)^k\right\} \\
&\leq \frac{1}{2\ln 2}\sum_{p=1}^{\infty}\frac{1}{p(2p-1)} = 1.
\end{aligned}$$

Since $B < 1$, the LHS of (2.C.4) is monotonically decreasing in ε. We therefore deduce that the inequality (2.C.4) holds for $\varepsilon \geq \varepsilon_0$, where ε_0 is the solution of

$$(1-C)B - \varepsilon_0 C(1-B) = 0.$$

It can be readily seen that the solution of the last equation is given by ε_0 defined in (2.82).

2.C.4 Proof of Eq. (2.83)

We will show both that there exists a unique ε_0 that satisfies (2.83), and that the RHS of (2.81) is non-positive if and only if $\varepsilon \geq \varepsilon_0$ where ε_0 is that unique solution. As in Appendix 2.C.3, we begin by multiplying the two sides of (2.81) by R and denoting $R = (1-\varepsilon)C$. It follows that the bound in the RHS of (2.81) is trivial (non-positive) if and only if

$$-\varepsilon C + \frac{1-(1-\varepsilon)C}{2\ln 2} \sum_{p=1}^{\infty} \frac{g_p^{\frac{(2-(1-\varepsilon)C)t}{1-(1-\varepsilon)C}}}{p(2p-1)} \leq 0 \tag{2.C.5}$$

where g_p is introduced in (2.57). We now show that the LHS of the last inequality is monotonically decreasing in ε. Let us denote

$$f(\varepsilon) \triangleq -\varepsilon C + \frac{1-(1-\varepsilon)C}{2\ln 2} \sum_{p=1}^{\infty} \frac{g_p^{\frac{(2-(1-\varepsilon)C)t}{1-(1-\varepsilon)C}}}{p(2p-1)}$$

$$\alpha_p \triangleq \frac{1}{2\ln 2\, p(2p-1)}.$$

By Dividing the derivative of f w.r.t. ε by C, we get

$$\begin{aligned}
\frac{f'(\varepsilon)}{C} &= \frac{1}{C}\Bigg(-C + C\sum_{p=1}^{\infty} \alpha_p\, g_p^{\frac{(2-(1-\varepsilon)C)t}{1-(1-\varepsilon)C}} \\
&\qquad + \big(1-(1-\varepsilon)C\big) \sum_{p=1}^{\infty} \alpha_p\, g_p^{\frac{(2-(1-\varepsilon)C)t}{1-(1-\varepsilon)C}} \log(g_p) \left(-\frac{tC}{(1-(1-\varepsilon)C)^2} \right) \Bigg) \\
&= \sum_{p=1}^{\infty} \left\{ \alpha_p \left(1 - \log\left(g_p^{\frac{t}{1-(1-\varepsilon)C}} \right) \right) g_p^{\frac{(2-(1-\varepsilon)C)t}{1-(1-\varepsilon)C}} \right\} - 1.
\end{aligned} \tag{2.C.6}$$

From the symmetry property of the pdf a then (2.57) yields that

$$\begin{aligned}
g_p &= \int_0^{\infty} a(l)(1+e^{-l}) \tanh^{2p}\left(\frac{l}{2}\right) \mathrm{d}l \\
&= \int_{-\infty}^{\infty} a(l) \tanh^{2p}\left(\frac{l}{2}\right) \mathrm{d}l \\
&\leq \int_{-\infty}^{\infty} a(l)\mathrm{d}l = 1
\end{aligned}$$

and since $g_p \leq 1$, then it follows that $g_p^{\frac{(2-(1-\varepsilon)C)t}{1-(1-\varepsilon)C}} \leq g_p^{\frac{t}{1-(1-\varepsilon)C}}$. Therefore, (2.C.6) gives

$$\frac{f'(\varepsilon)}{C} \leq \sum_{p=1}^{\infty} \left\{ \alpha_p \left(1 - \log\big(g_p^{\frac{t}{1-(1-\varepsilon)C}}\big) \right) g_p^{\frac{t}{1-(1-\varepsilon)C}} \right\} - 1.$$

For $p \in \mathbb{N}$, let us denote $g_p^{\frac{t}{1-(1-\varepsilon)C}} \triangleq 1 - \delta_p$ where $0 < \delta_p < 1$, then the last inequality gives

$$\begin{aligned} \frac{f'(\varepsilon)}{C} &\leq \sum_{p=1}^{\infty} \Big\{ \alpha_p \big(1 - \log(1-\delta_p)\big)(1-\delta_p) \Big\} - 1 \\ &\leq \sum_{p=1}^{\infty} \alpha_p - 1 = 0 \end{aligned}$$

where the second transition follows from the inequality $\ln(1-x) > -\frac{x}{1-x}$ which holds for $x \in (0,1)$. This concludes the proof of the monotonicity of the LHS of (2.C.5). Observing that

$$f(0) = (1-C) \sum_{p=1}^{\infty} \alpha_p \, g_p^{\frac{(2-C)t}{1-C}} > 0$$

and

$$\begin{aligned} f(1) &= -C + \sum_{p=1}^{\infty} \alpha_p \, g_p^{2t} \\ &\leq -C + \sum_{p=1}^{\infty} \alpha_p \, g_p \\ &= -C + C = 0 \end{aligned}$$

where the first inequality follows since $g_p \leq 1$ and since $t \geq 1$ ($t = 1$ if and only if the code is cycle-free, otherwise $t > 1$.) The second equality follows from the last three equalities leading to (2.75). From the continuity of the function f w.r.t. ε, we conclude that the monotonicity property of f, as shown above, ensures a unique solution for (2.83). From (2.C.5), it also follows from the monotonicity and continuity properties of the function f in terms of $\varepsilon \in (0,1)$ that the RHS of (2.81) is non-positive if and only if $\varepsilon \geq \varepsilon_0$ where ε_0 is the unique solution of (2.83).

Chapter 3

On Achievable Rates and Complexity of LDPC Codes over Parallel Channels: Bounds and Applications

This chapter is a reprint of

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

Chapter Overview: A variety of communication scenarios can be modeled by a set of parallel channels. The paper presents upper bounds on the achievable rates under maximum-likelihood decoding, and lower bounds on the decoding complexity per iteration of ensembles of low-density parity-check (LDPC) codes. The communication of these codes is assumed to take place over statistically independent parallel channels where the component channels are memoryless, binary-input and output-symmetric. The bounds are applied to ensembles of punctured LDPC codes where the puncturing patterns are either random or possess some structure. A diagram concludes our discussion by showing interconnections between the new theorems and some previously reported results.

3.1 Introduction

Parallel channels serve as a model for analyzing various communication scenarios, e.g., rate-compatible puncturing of error-correcting codes, non-uniformly error-protected codes, transmission over block-fading channels and multi-carrier signaling. All these scenarios can be modeled as a transmission of information over a set of parallel channels where each code symbol is assigned to one of these component channels. Naturally, analytical tools for evaluating the performance and decoding complexity of error-correcting codes whose transmission takes place over a set of parallel channels are gaining theoretical and practical interest (see, e.g., [49, 50, 77]).

The channel model considered in this paper assumes that the communication of binary linear block codes takes place over J statistically independent component channels where each of the individual channels is a memoryless binary-input output-symmetric (MBIOS) channel whose probability density function is given by $p(\cdot|\cdot\,;j)$ $(j = 1, 2, \ldots, J)$. If we let $\mathcal{I}(j)$ denote the set of indices of the symbols in an n-length codeword which are transmitted over the j^{th} channel, then

$$p_n(\mathbf{y}|\mathbf{x}) = \prod_{j=1}^{J} \prod_{i \in \mathcal{I}(j)} p(y_i|x_i;j)\,. \tag{3.1}$$

This paper focuses primarily on information-theoretic aspects of low-density parity-check (LDPC) codes whose transmission takes place over a set of parallel channels. It provides upper bounds on the achievable rates under maximum-likelihood (ML) decoding, and lower bounds on the decoding complexity per iteration of ensembles of LDPC codes. The paper forms a generalization of the results in [118]. However, the bounds on the achievable rates and decoding complexity derived in this paper are valid in probability 1 for ensembles of LDPC codes as one lets their block length tend to infinity; this is in contrast to the results in [118] which refer to communication over a single MBIOS channel and are valid code by code. The bounds introduced in this paper are applied to ensembles of punctured LDPC codes where the puncturing patterns are either random or possess some structure.

The performance of punctured LDPC codes under ML decoding was studied in [39] via analyzing the asymptotic growth rate of their average weight distributions and using upper bounds on the decoding error probability under ML decoding. Based on this analysis, it was proved that for any MBIOS channel, capacity-achieving codes of any desired rate can be constructed by puncturing the code bits of ensembles of LDPC codes whose design rate (before puncturing) is sufficiently low. The performance of punctured LDPC codes over the AWGN channel was studied in [35] under iterative message-passing decoding. Ha et al. studied in [35] two methods for puncturing

LDPC codes where the first method assumes random puncturing of the code bits at a fixed rate, and the second method assumes possibly different puncturing rates for each subset of code bits which corresponds to variable nodes of a fixed degree. For the second approach, called 'intentional puncturing', the degree distributions of the puncturing patterns were optimized in [34, 35] where it was aimed to minimize the threshold under iterative decoding for a given design rate via the Gaussian approximation; exact values of these optimized puncturing patterns were also calculated by the density evolution analysis and show good agreement with the Gaussian approximation. The results in [34, 35] exemplify the usefulness of punctured LDPC codes for a relatively wide range of rates, and therefore, they are suitable for rate-compatible coding.

The transmission of punctured codes over a single channel can be regarded as a special case of communication of the original code over a set of parallel channels (where this set of parallel channels is defined by the puncturing rates applied to subsets of the code bits). We therefore apply the bounds on the achievable rates and decoding complexity of LDPC codes over statistically independent parallel channels to the case of transmission of ensembles of punctured LDPC codes over a single MBIOS channel. We state puncturing theorems related to achievable rates and decoding complexity of punctured LDPC codes. For ensembles of punctured LDPC codes, the calculation of bounds on their thresholds under ML decoding and their exact thresholds under iterative decoding (based on density evolution analysis) is of interest in the sense that it enables one to separate the loss due to iterative decoding from the loss due to the structure of the ensembles.

The paper is organized as follows: Section 3.2 derives bounds on the conditional entropy of the transmitted codeword given the received sequence at the output of the parallel channels where the component channels are considered to be MBIOS. Section 3.3 relies on the previous bounds and derives an upper bound on the achievable rates of LDPC codes under ML decoding for parallel channels. Section 3.4 uses the latter result for the derivation of upper bounds on the achievable rates of ensembles of randomly and intentionally punctured LDPC codes whose transmission takes place over MBIOS channels, and numerical results are exemplified for various ensembles. Section 3.5 provides a lower bound on the decoding complexity (per iteration) of ensembles of LDPC codes under iterative message-passing decoding for parallel MBIOS channels. The latter result is used for the derivation of lower bounds on the decoding complexity of randomly and intentionally punctured LDPC codes for MBIOS channels; looser versions of these bounds suggest a simplified re-derivation of previously reported bounds on the decoding complexity of randomly punctured LDPC codes (as

shown in an appendix). Finally, Section 3.6 summarizes our discussion, and presents a diagram which shows interconnections between the theorems introduced in this paper and some other previously reported results from [17, 65, 69, 67, 81, 118]. The preliminary material on ensembles of LDPC codes and notation required for this paper are introduced in [74] and [118, Section 2].

3.2 Bounds on the Conditional Entropy for Parallel Channels

This section serves as a preparatory step towards the derivation of upper bounds on the achievable rates of ML decoded binary linear block codes whose transmission takes place over statistically independent parallel MBIOS channels. To this end, we present in this section upper and lower bounds on the conditional entropy of the transmitted codeword given the received sequence at the output of these channels.

3.2.1 Lower Bound on the Conditional Entropy

We begin by deriving an information-theoretic lower bound on the conditional entropy of the transmitted codeword given the received sequence, when the transmission takes place over a set of J independent parallel MBIOS channels.

Proposition 3.1 Let $\mathcal{C}$ be a binary linear block code of length n, and assume that its transmission takes place over a set of J statistically independent parallel MBIOS channels. Let C_j denote the capacity of the j^{th} channel (in bits per channel use), and $a(\cdot; j)$ designate the conditional pdf of the log-likelihood ratio (LLR) at the output of the j^{th} channel given its input is 0. Let $\mathbf{X} = (X_1, \ldots X_n)$ and $\mathbf{Y} = (Y_1, \ldots, Y_n)$ designate the transmitted codeword and received sequence, respectively, $\mathcal{I}(j)$ be the set of indices of the code bits transmitted over the j^{th} channel, $n^{[j]} \triangleq |\mathcal{I}(j)|$ be the size of this set, and $p_j \triangleq \frac{n^{[j]}}{n}$ be the fraction of bits transmitted over the j^{th} channel. For an arbitrary $c \times n$ parity-check matrix H of the code $\mathcal{C}$, let $\beta_{j,m}$ designate the number of indices in $\mathcal{I}(j)$ referring to bits which are involved in the m^{th} parity-check equation of H (where $m \in \{1, \ldots, c\}$), and let $R_{\text{d}} = 1 - \frac{c}{n}$ be the design rate of $\mathcal{C}$. Then, the conditional entropy of the transmitted codeword given the received sequence satisfies

$$\frac{H(\mathbf{X}|\mathbf{Y})}{n} \geq 1 - \sum_{j=1}^{J} p_j C_j - (1 - R_{\text{d}}) \left(1 - \frac{1}{2n(1-R_{\text{d}}) \ln 2} \sum_{p=1}^{\infty} \left\{ \frac{1}{p(2p-1)} \sum_{m=1}^{n(1-R_{\text{d}})} \prod_{j=1}^{J} (g_{j,p})^{\beta_{j,m}} \right\} \right) \tag{3.2}$$

where

$$g_{j,p} \triangleq \int_0^\infty a(l;j)\,(1+e^{-l})\tanh^{2p}\left(\frac{l}{2}\right)\mathrm{d}l, \qquad j \in \{1,\ldots,J\}, \quad p \in \mathbb{N}. \tag{3.3}$$

Remark 3.1 Note that the input vector $\mathbf{X}$ is chosen uniformly from the codewords of a binary linear block code. Each input bit X_i therefore either gets the values 0 or 1 with probability $\frac{1}{2}$ or is set to zero (due to the linearity of the code). In the following proof, we assume that all the code symbols get the values 0 or 1 with equal probability. By slightly modifying the proof, it is simple to show that the bound also holds for the other case where some of the code bits are set to zero.

Proof: The proof relies on concepts which are presented in [17, 118], and generalizes them to the case of parallel channels. If a symbol is transmitted over the j^{th} MBIOS channel and y is the corresponding output, then the LLR gets the form

$$\mathrm{LLR}(y;j) = \ln\left(\frac{p(y|0;j)}{p(y|1;j)}\right), \qquad j \in \{1,\ldots,J\}, \quad y \in \mathcal{Y}$$

where $\mathcal{Y}$ denotes the output alphabet of each component channel,[1] and $p(\cdot|\cdot;j)$ is the conditional pdf of the j^{th} channel. For each one of these J component channels, we move from the original mapping of $X \to Y$ (where according to (3.1), each symbol is transmitted over only one of these J channels) to an equivalent representation of the channel $X \to \widetilde{Y}$, where $\widetilde{Y}$ represents the LLR of the channel output Y. These channels are equivalent in the sense that $H(X|\widetilde{Y}) = H(X|Y)$. The basic idea for showing the equivalence between the original set of parallel channels and the one which will be introduced shortly is based on the principle that the LLR forms a sufficient statistics of an MBIOS channel.

In the following, we characterize an equivalent channel to each of the J parallel channels. The output of the equivalent channel is defined to be $\widetilde{Y} \triangleq (\Phi, \Omega)$. For the j-th channel, $\widetilde{Y}$ is calculated from Y as follows:

$$\Omega \triangleq |\mathrm{LLR}(Y;j)|, \quad \Phi \triangleq \begin{cases} 0 & \text{if} \quad \mathrm{LLR}(Y;j) > 0 \\ 1 & \text{if} \quad \mathrm{LLR}(Y;j) < 0 \\ 0 \text{ or } 1 \text{ w.p. } \frac{1}{2} & \text{if} \quad \mathrm{LLR}(Y;j) = 0 \end{cases}.$$

Due to the symmetry of the communication channel, the equivalent channel can be seen as a channel with additive noise where the transmitted signal affects only the Φ component of the output $\widetilde{Y}$. The characterization of the equivalent channel in this

[1] In case the output alphabets of the component channels are not equal, then $\mathcal{Y}$ can be defined as their union.

form is used for the continuation of this proof and is presented below. For each index $i \in \mathcal{I}(j)$, let us choose independently a value L_i according to the conditional pdf $a(\cdot; j)$, and for $i \in \{1, \ldots, n\}$, let

$$\Omega_i \triangleq |L_i|, \quad \Theta_i \triangleq \begin{cases} 0 & \text{if} \quad L_i > 0 \\ 1 & \text{if} \quad L_i < 0 \\ 0 \text{ or } 1 \text{ w.p. } \frac{1}{2} & \text{if} \quad L_i = 0 \end{cases}.$$

The output of the set of equivalent channels is defined as $\widetilde{\mathbf{Y}} = (\widetilde{Y}_1, \ldots, \widetilde{Y}_n)$ where $\widetilde{Y}_i = (\Phi_i, \Omega_i)$ and $\Phi_i = \Theta_i + X_i$ where the addition is modulo 2. This defines the mapping

$$X \to \widetilde{Y} = (\Phi, \Omega)$$

where Φ is a binary random variable which is affected by X, and Ω is a non-negative random variable which is independent of X. Note that due to the symmetry of the parallel channels, for each index $i \in \mathcal{I}(j)$, the joint distribution of (Φ_i, Ω_i) is independent of i, and is equal to the distribution of the pair representing the sign and magnitude of $\mathrm{LLR}(Y; j)$. Hence,

$$f_{\Omega_i}(\omega) \triangleq f_{\Omega}(\omega; j) = \begin{cases} a(\omega; j) + a(-\omega; j) = (1 + e^{-\omega})\, a(\omega; j) & \text{if} \quad \omega > 0 \\ a(0; j) & \text{if} \quad \omega = 0 \end{cases} \tag{3.4}$$

where we rely on the symmetry property of $a(\cdot; j)$.

Denoting by R the rate of the code $\mathcal{C}$, since the codewords are transmitted with equal probability

$$H(\mathbf{X}) = nR. \tag{3.5}$$

Also, since the J parallel channels are memoryless, then

$$H(\mathbf{Y}|\mathbf{X}) = \sum_{i=1}^{n} H(Y_i|X_i). \tag{3.6}$$

The mapping $Y_i \to \widetilde{Y}_i$ is memoryless, hence $H(\widetilde{\mathbf{Y}}|\mathbf{Y}) = \sum_{i=1}^{n} H(\widetilde{Y}_i|Y_i)$, and

$$\begin{aligned} H(\mathbf{Y}) &= H(\widetilde{\mathbf{Y}}) - H(\widetilde{\mathbf{Y}}|\mathbf{Y}) + H(\mathbf{Y}|\widetilde{\mathbf{Y}}) \\ &= H(\widetilde{\mathbf{Y}}) - \sum_{i=1}^{n} H(\widetilde{Y}_i|Y_i) + H(\mathbf{Y}|\widetilde{\mathbf{Y}}) \end{aligned} \tag{3.7}$$

$$\begin{aligned} H(\mathbf{Y}|\widetilde{\mathbf{Y}}) &\leq \sum_{i=1}^{n} H(Y_i|\widetilde{Y}_i) \\ &= \sum_{i=1}^{n} \Big[H(Y_i) - H(\widetilde{Y}_i) + H(\widetilde{Y}_i|Y_i) \Big]. \end{aligned} \tag{3.8}$$

Applying the above towards the derivation of a lower bound on the conditional entropy $H(\mathbf{X}|\mathbf{Y})$, we get

$$\begin{aligned}
H(\mathbf{X}|\mathbf{Y}) &= H(\mathbf{X}) + H(\mathbf{Y}|\mathbf{X}) - H(\mathbf{Y}) \\
&\overset{(a)}{=} nR + \sum_{i=1}^{n} H(Y_i|X_i) - H(\widetilde{\mathbf{Y}}) - H(\mathbf{Y}|\widetilde{\mathbf{Y}}) + \sum_{i=1}^{n} H(\widetilde{Y}_i|Y_i) \\
&\overset{(b)}{\geq} nR + \sum_{i=1}^{n} H(Y_i|X_i) - H(\widetilde{\mathbf{Y}}) - \sum_{i=1}^{n} \Big[H(Y_i) - H(\widetilde{Y}_i) + H(\widetilde{Y}_i|Y_i) \Big] \\
&\quad + \sum_{i=1}^{n} H(\widetilde{Y}_i|Y_i) \\
&= nR - H(\widetilde{\mathbf{Y}}) + \sum_{i=1}^{n} H(\widetilde{Y}_i) - \sum_{i=1}^{n} \big[H(Y_i) - H(Y_i|X_i) \big] \\
&= nR - H(\widetilde{\mathbf{Y}}) + \sum_{i=1}^{n} H(\widetilde{Y}_i) - \sum_{i=1}^{n} I(X_i;Y_i) \\
&\overset{(c)}{\geq} nR - H(\widetilde{\mathbf{Y}}) + \sum_{i=1}^{n} H(\widetilde{Y}_i) - \sum_{j=1}^{J} n^{[j]}\, C_j \qquad (3.9)
\end{aligned}$$

where (a) relies on (3.5)–(3.7), (b) relies on (3.8), and (c) follows since $I(X_i;Y_i) \leq C_j$ for all $i \in \mathcal{I}(j)$, and $|\mathcal{I}(j)| = n^{[j]}$ for $j \in \{1,\ldots,J\}$. In order to obtain a lower bound on $H(\mathbf{X}|\mathbf{Y})$ from (3.9), we calculate the entropy of the random variables $\{\widetilde{Y}_i\}$, and find an upper bound on the entropy of the random vector $\widetilde{\mathbf{Y}}$. This finally provides the lower bound on the conditional entropy given in (3.2). Considering an index $i \in \mathcal{I}(j)$ for some $j \in \{1,2,\ldots J\}$, we get

$$\begin{aligned}
H(\widetilde{Y}_i) &= H(\Phi_i, \Omega_i) \\
&= H(\Omega_i) + H(\Phi_i|\Omega_i) \\
&= H(\Omega_i) + \mathbb{E}_{\omega}\left[H(\Phi_i|\Omega_i = \omega)\right] \\
&= H(\Omega_i) + 1 \qquad (3.10)
\end{aligned}$$

where the last transition is due to the fact that given the absolute value of the LLR, since the parallel channels are MBIOS and the coded bits are equally likely to be 0 or 1, the sign of the LLR is equally likely to be positive or negative. The entropy $H(\Omega_i)$ is not expressed explicitly as it cancels out later.

We now turn to derive an upper bound on $H(\widetilde{\mathbf{Y}})$:

$$\begin{aligned}
H(\widetilde{\mathbf{Y}}) &= H\big((\Phi_1,\ldots,\Phi_n),\,(\Omega_1,\ldots,\Omega_n)\big) \\
&= H(\Omega_1,\ldots,\Omega_n) + H\big((\Phi_1,\ldots,\Phi_n)\,|\,(\Omega_1,\ldots,\Omega_n)\big) \\
&= \sum_{i=1}^{n} H(\Omega_i) + H\big((\Phi_1,\ldots,\Phi_n)\,|\,(\Omega_1,\ldots,\Omega_n)\big) \qquad (3.11)
\end{aligned}$$

where the last equality follows since the random variables Ω_i are statistically independent.

Define the c-dimensional syndrome vector as

$$\mathbf{S} \triangleq (\Phi_1, \ldots, \Phi_n)\, H^T$$

where H is a $c \times n$ parity-check matrix of the binary linear block code $\mathcal{C}$, and let L be the index of the vector $(\Phi_1, \ldots, \Phi_n)$ in the coset which corresponds to $\mathbf{S}$. Since each coset has exactly 2^{nR} elements which are equally likely then $H(L) = nR$, and we get

$$\begin{aligned} H\big((\Phi_1, \ldots, \Phi_n) \,|\, (\Omega_1, \ldots, \Omega_n)\big) &= H\big(\mathbf{S}, L \,|\, (\Omega_1, \ldots, \Omega_n)\big) \\ &\leq H(L) + H\big(\mathbf{S} \,|\, (\Omega_1, \ldots, \Omega_n)\big) \\ &= nR + H\big(\mathbf{S} \,|\, (\Omega_1, \ldots, \Omega_n)\big) \\ &\leq nR + \sum_{m=1}^{c} H\big(S_m \,|\, (\Omega_1, \ldots, \Omega_n)\big). \end{aligned} \tag{3.12}$$

Since $\mathbf{X}\, H^T = 0$ for any codeword, then

$$\mathbf{S} = (\Theta_1, \ldots, \Theta_n)\, H^T.$$

Let us consider the m^{th} parity-check equation which involves k_m variables, and assume that the set of indices of these variables is $\{i_1, \ldots, i_{k_m}\}$. Then, the component S_m of the syndrome is equal to 1 if and only if there is an odd number of ones in the random vector $(\Theta_{i_1}, \ldots, \Theta_{i_{k_m}})$. To calculate the probability that S_m is equal to 1, we rely on the following lemma:

Lemma 3.1 ([118], Lemma 4.1) If the m^{th} linear constraint defined by the parity-check matrix H involves k_m variables, and if $\{i_1, \ldots, i_{k_m}\}$ denote the indices of these variables, then

$$\Pr\left(S_m = 1 \,|\, (\Omega_{i_1}, \ldots, \Omega_{i_{k_m}}) = (\alpha_1, \ldots, \alpha_{k_m})\right) = \frac{1}{2}\left[1 - \prod_{w=1}^{k_m} \tanh\left(\frac{\alpha_w}{2}\right)\right]. \tag{3.13}$$

From this lemma, we obtain

$$H\left(S_m | (\Omega_{i_1}, \ldots, \Omega_{i_{k_m}}) = (\alpha_1, \ldots, \alpha_{k_m})\right) = h_2\left(\frac{1}{2}\left[1 - \prod_{w=1}^{k_m} \tanh\left(\frac{\alpha_w}{2}\right)\right]\right)$$

where h_2 denotes the binary entropy function to base 2. By taking the statistical expectation over the k_m random variables $\Omega_{i_1}, \ldots, \Omega_{i_{k_m}}$, we get

$$\begin{aligned} &H\left(S_m | (\Omega_{i_1}, \ldots, \Omega_{i_{k_m}})\right) \\ &= \int_0^\infty \ldots \int_0^\infty h_2\left(\frac{1}{2}\left[1 - \prod_{w=1}^{k_m} \tanh\left(\frac{\alpha_w}{2}\right)\right]\right) \prod_{w=1}^{k_m} f_{\Omega_{i_w}}(\alpha_w)\, \mathrm{d}\alpha_1 \mathrm{d}\alpha_2 \ldots \mathrm{d}\alpha_{k_m}\,. \end{aligned}$$

Let $\beta_{j,m}$ denote the number of indices $w \in \{i_1, \ldots, i_{k_m}\}$ referring to variables which are transmitted over the j^{th} channel. From the Taylor series expansion of the binary entropy function (h_2) around $x = \frac{1}{2}$ (see [118, Appendix B.1])

$$h_2(x) = 1 - \frac{1}{2\ln 2} \sum_{p=1}^{\infty} \frac{(1-2x)^{2p}}{p(2p-1)}, \quad 0 \leq x \leq 1 \tag{3.14}$$

it follows that

$$\begin{aligned} & H\left(S_m | (\Omega_{i_1}, \ldots, \Omega_{i_{k_m}})\right) \\ &= 1 - \frac{1}{2\ln 2} \sum_{p=1}^{\infty} \left\{ \frac{1}{p(2p-1)} \prod_{w=1}^{k_m} \left(\int_0^{\infty} f_{\Omega_{i_w}}(\alpha) \, \tanh^{2p}\left(\frac{\alpha}{2}\right) \, \mathrm{d}\alpha \right) \right\} \\ &= 1 - \frac{1}{2\ln 2} \sum_{p=1}^{\infty} \left\{ \frac{1}{p(2p-1)} \prod_{j=1}^{J} \left(\int_0^{\infty} f_{\Omega}(\alpha; j) \, \tanh^{2p}\left(\frac{\alpha}{2}\right) \mathrm{d}\alpha \right)^{\beta_{j,m}} \right\} \end{aligned} \tag{3.15}$$

where the first transition is based on (3.14) and follows along the same lines as [118, Appendix B.2]), and the second transition is due to the fact that for all $i \in \mathcal{I}(j)$, the pdf of the random variable Ω_i is independent of i, see (3.4). Summing over all the parity-check equations of H gives

$$\begin{aligned} & \sum_{m=1}^{c} H\left(S_m | (\Omega_1, \ldots, \Omega_n)\right) \\ &= c - \frac{1}{2\ln 2} \sum_{p=1}^{\infty} \frac{1}{p(2p-1)} \sum_{m=1}^{c} \left\{ \prod_{j=1}^{J} \left(\int_0^{\infty} f_{\Omega}(\alpha; j) \tanh^{2p}\left(\frac{\alpha}{2}\right) \mathrm{d}\alpha \right)^{\beta_{j,m}} \right\}. \end{aligned} \tag{3.16}$$

By combining (3.4), (3.11), (3.12) and (3.16), we get the following upper bound on $H(\widetilde{\mathbf{Y}})$:

$$\begin{aligned} H(\widetilde{\mathbf{Y}}) \quad &\leq \quad \sum_{i=1}^{n} H(\Omega_i) + nR + c \left[1 - \frac{1}{2c\ln 2} \sum_{p=1}^{\infty} \frac{1}{p(2p-1)} \right. \\ &\qquad \left. \cdot \sum_{m=1}^{c} \left\{ \prod_{j=1}^{J} \left(\int_0^{\infty} a(\alpha; j)(1+e^{-\alpha}) \tanh^{2p}\left(\frac{\alpha}{2}\right) \mathrm{d}\alpha \right)^{\beta_{j,m}} \right\} \right] \\ &\overset{(a)}{=} \quad \sum_{i=1}^{n} H(\Omega_i) + nR + n(1 - R_{\mathrm{d}}) \\ &\qquad \cdot \left[1 - \frac{1}{2n(1-R_{\mathrm{d}})\ln 2} \sum_{p=1}^{\infty} \left\{ \frac{1}{p(2p-1)} \sum_{m=1}^{c} \prod_{j=1}^{J} {g_{j,p}}^{\beta_{j,m}} \right\} \right] \end{aligned} \tag{3.17}$$

where (a) relies on the definition of $g_{j,p}$ in (3.3) and since $R_{\mathrm{d}} \triangleq 1 - \frac{c}{n}$ denotes the design rate of $\mathcal{C}$. Finally, the substitution of (3.10) and (3.17) in the RHS of (3.9)

provides the lower bound on the conditional entropy $H(\mathbf{X}|\mathbf{Y})$ given in (3.2). This completes the proof of the proposition. ■

3.2.2 Upper Bound on the Conditional Entropy

In this section, we provide an upper bound on the conditional entropy of the transmitted codeword given the received sequence. The bound holds for an arbitrary binary linear block code whose transmission takes place over a set of parallel channels, and is expressed in terms of the code rate and the bit-error probability of the code (under ML decoding or a sub-optimal decoding algorithm).

Lemma 3.2 Let $\mathcal{C}$ be a binary linear block code of length n and rate R, and assume that its transmission takes place over a set of parallel channels. Let $\mathbf{X} = (X_1, \ldots, X_n)$ and $\mathbf{Y} = (Y_1, \ldots, Y_n)$ designate the transmitted codeword and the received sequence, respectively. Then

$$\frac{H(\mathbf{X}|\mathbf{Y})}{n} \leq R\, h_2(P_{\mathrm{b}}) \tag{3.18}$$

where P_{b} designates the bit error probability of the code $\mathcal{C}$ under an arbitrary decoding algorithm.

Proof: Since there is a one to one correspondence between the codewords and the set of information bits used to encode them, then $H(\mathbf{X}|\mathbf{Y}) = H(\mathbf{U}|\mathbf{Y})$ where the vector $\mathbf{U} = (U_1, \ldots, U_{nR})$ denotes the sequence of information bits used to encode the codeword $\mathbf{X}$. Let $P_{\mathrm{b}}^{(i)}(\mathcal{C})$ denote the probability of decoding the bit U_i erroneously given the received sequence at the output of the set of parallel channels, then the bit error probability is given by

$$P_{\mathrm{b}}(\mathcal{C}) = \frac{1}{nR} \sum_{i=1}^{nR} P_{\mathrm{b}}^{(i)}(\mathcal{C}). \tag{3.19}$$

This therefore gives

$$\begin{aligned}
\frac{H(\mathbf{X}|\mathbf{Y})}{n} &= \frac{H(\mathbf{U}|\mathbf{Y})}{n} \\
&\overset{(a)}{\leq} \frac{1}{n} \sum_{i=1}^{nR} H(U_i|\mathbf{Y}) \\
&\overset{(b)}{\leq} \frac{1}{n} \sum_{i=1}^{nR} h_2\big(P_{\mathrm{b}}^{(i)}(\mathcal{C})\big) \\
&\overset{(c)}{\leq} R\, h_2\left(\frac{1}{nR} \sum_{i=1}^{nR} P_{\mathrm{b}}^{(i)}(\mathcal{C})\right) \\
&\overset{(d)}{=} R\, h_2\big(P_{\mathrm{b}}(\mathcal{C})\big)
\end{aligned}$$

where inequality (a) holds from the chain rule of entropy and since conditioning reduces entropy, inequality (b) follows from Fano's inequality and since the code is binary, inequality (c) is based on Jensen's inequality and the concavity of the binary entropy function, and equality (d) follows from (3.19). ■

3.3 An Upper Bound on the Achievable Rates of LDPC codes over Parallel Channels

In this section, we derive an upper bound on the design rate of a sequence of ensembles of LDPC codes whose transmission takes place over a set of statistically independent parallel MBIOS channels, and which achieves vanishing bit error probability under ML decoding. This bound is used in the next section for the derivation of an upper bound on the design rate of an arbitrary sequence of ensembles of punctured LDPC codes.

Let us assume that a binary LDPC code $\mathcal{C}$ of length n is transmitted over a set of J statistically independent parallel MBIOS channels. Denote the number of code bits of $\mathcal{C}$ which are transmitted over the j^{th} channel by $n^{[j]}$, and the fraction of bits transmitted over the j^{th} channel by

$$p_j \triangleq \frac{n^{[j]}}{n}, \quad j \in \{1, \ldots, J\}. \tag{3.20}$$

Let $\mathcal{G}$ be a bipartite graph which represents the code $\mathcal{C}$, and E be the set of edges in $\mathcal{G}$. Let $E^{[j]}$ designate the set of edges connected to variable nodes which correspond to code bits transmitted over the j^{th} channel, and

$$q_j \triangleq \frac{|E^{[j]}|}{|E|}, \quad j \in \{1, \ldots, J\} \tag{3.21}$$

denote the fraction of edges connected to these variable nodes. Referring to the edges from the subset $E^{[j]}$, let $\lambda_i^{[j]}$ designate the fraction of these edges which are connected to variable nodes of degree i, and define the following J degree distributions from the edge perspective:

$$\lambda^{[j]}(x) \triangleq \sum_{i=2}^{\infty} \lambda_i^{[j]} x^{i-1}, \qquad j \in \{1, \ldots, J\}$$

which correspond to each of the J parallel channels. According to this notation, the number of edges connected to variable nodes corresponding to code bits transmitted

over the j^{th} channel is given by

$$|E^{[j]}| = \frac{n^{[j]}}{\sum_{i=2}^{\infty} \frac{\lambda_i^{[j]}}{i}}, \qquad j \in \{1, \ldots, J\}. \tag{3.22}$$

For the simplicity of the notation, let us define a vector of degree distributions for the variable nodes from the edge perspective to be $\lambda(x) = \big(\lambda^{[1]}(x), \ldots, \lambda^{[J]}(x)\big)$. Following the notation in [69], the ensemble (n, λ, ρ) is defined to be the set of LDPC codes of length n, which according to their representation by bipartite graphs and the assignment of their code bits to the parallel channels, imply left and right degree distributions of λ and ρ, respectively.

Lemma 3.3

$$\frac{1}{\int_0^1 \lambda(x)\,\mathrm{d}x} = \sum_{j=1}^{J} \left\{ \frac{p_j}{\int_0^1 \lambda^{[j]}(x)\,\mathrm{d}x} \right\}. \tag{3.23}$$

where λ is the overall left degree distribution which serves to construct the vector of left degree distributions λ by considering the assignments of variables nodes to the different channels.

Proof: Since $E^{[1]}, \ldots, E^{[J]}$ forms a sequence of disjoint sets whose union is the set E, we get the equality $|E| = \sum_{j=1}^{J} |E^{[j]}|$. From (3.22), we therefore get

$$\frac{n}{\sum_{i=2}^{\infty} \frac{\lambda_i}{i}} = \sum_{j=1}^{J} \left\{ \frac{n^{[j]}}{\sum_{i=2}^{\infty} \frac{\lambda_i^{[j]}}{i}} \right\} \tag{3.24}$$

and by dividing both sides of the equality by n and using (3.20), the lemma follows immediately. ■

Lemma 3.4

$$q_j = \frac{p_j}{\int_0^1 \lambda^{[j]}(x)\,\mathrm{d}x} \cdot \frac{1}{\sum_{k=1}^{J} \left\{ \frac{p_k}{\int_0^1 \lambda^{[k]}(x)\,\mathrm{d}x} \right\}}, \quad \forall\, j \in \{1, \ldots, J\}. \tag{3.25}$$

Proof: The lemma follows directly from (3.21), (3.22) and Lemma 3.3. ■

In the following, we introduce a sequence of ensembles of LDPC codes, say $\{(n_r, \boldsymbol{\lambda}_{\mathbf{r}}, \rho)\}_{r=1}^{\infty}$ where all the codes in each ensemble have the same number of bits assigned to each of the J parallel channels, and ρ is fixed for all the ensembles of

this sequence (i.e., it is independent of r). Since λ which corresponds to the overall left degree distribution of the edges is also independent of r, one can consider here the common design rate of the sequence of ensembles $\{(n_r, \lambda_{\mathbf{r}}, \rho)\}_{r=1}^{\infty}$ which does not depend on r.

This setting is general enough for applying the following theorem to various applications which form particular cases of communication over parallel channels, e.g., punctured LDPC codes [35, 39], non-uniformly error protected LDPC codes [69], and LDPC-coded modulation (see e.g., [37, 112]). In this setting, the fraction of code bits assigned to the j^{th} channel, $p_{j,r}$, depends on $j \in \{1, \ldots, J\}$ and $r \in \mathbb{N}$, but not on the particular code chosen from each ensemble. It follows from Lemma 3.4 that the same property also holds for $q_{j,r}$ which designates the fraction of edges connected to variable nodes whose code bits are assigned to the j^{th} channel. In the following, we assume that the limits

$$p_j \triangleq \lim_{r\to\infty} p_{j,r}, \quad q_j \triangleq \lim_{r\to\infty} q_{j,r} \tag{3.26}$$

exist and we also assume that they are positive for all $j \in \{1, \ldots, J\}$ (though in general, they are non-negative).

Theorem 3.1 Let a sequence of LDPC ensembles $\{(n_r, \lambda_{\mathbf{r}}, \rho)\}_{r=1}^{\infty}$ be transmitted over a set of J statistically independent parallel MBIOS channels, and assume that the block length (n_r) goes to infinity as we let r tend to infinity. Let C_j denote the capacity of the j^{th} channel, and $a(\cdot; j)$ designate the pdf of the LLR at the output of the j^{th} channel given its input is 1. If in the limit where r tends to infinity, the bit error probability of this sequence under ML decoding vanishes, then the common design rate R_{d} of these ensembles satisfies

$$R_{\text{d}} \leq 1 - \frac{1 - \sum_{j=1}^{J} p_j C_j}{1 - \frac{1}{2\ln 2} \sum_{p=1}^{\infty} \left\{ \frac{1}{p(2p-1)} \Gamma\left(\sum_{j=1}^{J} q_j \, g_{j,p} \right) \right\}} \tag{3.27}$$

where Γ denotes the right degree distribution from the node perspective, and $g_{j,p}$ is introduced in (3.3).

Proof: Let $\{\mathcal{C}_r\}_{r=1}^{\infty}$ be a sequence of LDPC codes chosen uniformly at random from the sequence of ensembles $\{(n_r, \lambda_{\mathbf{r}}, \rho)\}_{r=1}^{\infty}$. Denote the rate of the code $\mathcal{C}_r$ by R_r, and let $P_{\text{b},r}$ be its bit error probability under ML decoding. Let $\mathcal{G}_r$ be a bipartite graph of the code $\mathcal{C}_r$ whose left and right degree distributions from the edge perspective are $\lambda_{\mathbf{r}}$ and ρ, respectively. From Proposition 3.1 and Lemma 3.2, it follows that the

following inequality holds for the binary linear block code $\mathcal{C}_r$:

$$R_r h_2(P_{\mathrm{b},r}) \geq 1 - \sum_{j=1}^{J} p_{j,r} C_j - (1 - R_{\mathrm{d}}) \cdot \left(1 - \frac{1}{2n_r(1-R_{\mathrm{d}})\ln 2} \sum_{p=1}^{\infty} \left\{ \frac{1}{p(2p-1)} \sum_{m=1}^{n_r(1-R_{\mathrm{d}})} \prod_{j=1}^{J} (g_{j,p})^{\beta_{r,j,m}} \right\} \right) \quad (3.28)$$

where n_r is the block length of the code $\mathcal{C}_r$, R_{d} is the common design rate for all the codes from the sequence of ensembles $\{(n_r, \lambda_{\mathbf{r}}, \rho)\}_{r=1}^{\infty}$, and $\beta_{r,j,m}$ denotes the number of edges which are connected to the m^{th} parity-check node of the graph $\mathcal{G}_r$ and are related to code bits transmitted over the j^{th} channel (where $j \in \{1, \ldots, J\}$ and $m \in \{1, \ldots n_r(1-R_{\mathrm{d}})\}$). By taking the expectation on both sides of (3.28) and letting r tend to infinity, we get

$$1 - \sum_{j=1}^{J} p_j C_j - (1 - R_{\mathrm{d}}) \lim_{r \to \infty} \left(1 - \frac{1}{2n_r(1-R_{\mathrm{d}})\ln 2} \cdot \sum_{p=1}^{\infty} \left\{ \frac{1}{p(2p-1)} \sum_{m=1}^{n_r(1-R_{\mathrm{d}})} \mathbb{E}\left(\prod_{j=1}^{J} (g_{j,p})^{\beta_{r,j,m}} \right) \right\} \right) \leq 0. \quad (3.29)$$

The RHS of (3.29) follows from the LHS of (3.28), due to the concavity of the binary entropy function and Jensen's inequality, and since by our assumption, the bit error probability of the ensembles vanishes in the limit where r tends to infinity.

The derivation of an upper bound on the design rate is proceeded by calculating the expectation of the product inside the LHS of (3.29). Let $k_{r,m}$ denote the degree of the m^{th} parity-check node of the bipartite graph $\mathcal{G}_r$, then the smoothing theorem gives

$$\mathbb{E}\left(\prod_{j=1}^{J} (g_{j,p})^{\beta_{r,j,m}} \right) = \mathbb{E}\left[\mathbb{E}\left(\prod_{j=1}^{J} (g_{j,p})^{\beta_{r,j,m}} \,\middle|\, \sum_{j=1}^{J} \beta_{r,j,m} = k_{r,m} \right) \right] \quad (3.30)$$

where the outer expectation is carried over the random variable $k_{r,m}$. We first calculate the inner expectation in the RHS of (3.30). It follows from (3.22) that the number of edges, $|E_r^{[j]}| \triangleq |E^{[j]}(\mathcal{G}_r)|$, connected to variable nodes corresponding to code bits transmitted over the j^{th} channel, is independent of the code $\mathcal{C}_r$ choosen from the ensemble $(n_r, \lambda_{\mathbf{r}}, \rho)$. The same property also holds for the total number of edges in the graph (since $|E_r| = \sum_{j=1}^{J} |E_r^{[j]}|$). Since the code $\mathcal{C}_r$ is chosen uniformly at random from the ensemble, it follows that if $k_{r,m}$ is a given positive integer, then

$$\mathbb{E}\left(\prod_{j=1}^{J}(g_{j,p})^{\beta_{r,j,m}} \,\Big|\, \sum_{j=1}^{J}\beta_{r,j,m}=k_{r,m}\right)$$

$$=\sum_{\substack{b_1,\ldots,b_J\geq 0\\ \sum_{j=1}^{J}b_j=k_{r,m}}}\left\{\Pr\big(\beta_{r,j,m}=b_j\,,\,\forall j\in\{1,\ldots,J\}\big)\prod_{j=1}^{J}(g_{j,p})^{b_j}\right\}$$

$$=\sum_{\substack{b_1,\ldots,b_J\geq 0\\ \sum_{j=1}^{J}b_j=k_{r,m}}}\left\{\frac{\binom{|E_r^{[1]}|}{b_1}\cdots\binom{|E_r^{[J]}|}{b_J}}{\binom{|E_r|}{k_{r,m}}}\prod_{j=1}^{J}(g_{j,p})^{b_j}\right\}. \tag{3.31}$$

Lemma 3.5

$$\lim_{r\to\infty}\frac{\binom{|E_r^{[1]}|}{b_1}\cdots\binom{|E_r^{[J]}|}{b_J}}{\binom{|E_r|}{k_{r,m}}}=\prod_{j=1}^{J}(q_j)^{b_j}\lim_{r\to\infty}\binom{k_{r,m}}{b_1,b_2,\ldots,b_J}. \tag{3.32}$$

Proof: By assumption, in the limit where we let r tend to infinity, the block length n_r also tends to infinity. Hence, from (3.22) and the assumption that $q_j>0$ for every $j\in\{1,\ldots,J\}$, we get that for all $j\in\{1,\ldots,J\}$, $E_r^{[j]}$ approaches infinity in the limit where r tends to infinity.

$$\lim_{r\to\infty}\frac{\binom{|E_r^{[1]}|}{b_1}\cdots\binom{|E_r^{[J]}|}{b_J}}{\binom{|E_r|}{k_{r,m}}}$$

$$=\lim_{r\to\infty}\frac{|E_r^{[1]}|!\cdots|E_r^{[J]}|!}{|E_r|!}\,\frac{(|E_r|-k_{r,m})!}{(|E_r^{[1]}|-b_1)!\cdots(|E_r^{[J]}|-b_J)!}\binom{k_{r,m}}{b_1,b_2,\ldots,b_J}$$

$$=\lim_{r\to\infty}\left\{\frac{|E_r^{[1]}|!\cdots|E_r^{[J]}|!}{|E_r|!}\,\frac{(|E_r|-k_{r,m})!}{(|E_r^{[1]}|-b_1)!\ldots(|E_r^{[J]}|-b_J)!}\right\}\lim_{r\to\infty}\binom{k_{r,m}}{b_1,b_2,\ldots,b_J}$$

$$\overset{(a)}{=}\lim_{r\to\infty}\frac{|E_r^{[1]}|^{b_1}\cdots|E_r^{[J]}|^{b_J}}{|E_r|^{k_{r,m}}}\lim_{r\to\infty}\binom{k_{r,m}}{b_1,b_2,\ldots,b_J}$$

$$\overset{(b)}{=}\lim_{r\to\infty}\left(\frac{|E_r^{[1]}|}{|E_r|}\right)^{b_1}\cdots\left(\frac{|E_r^{[J]}|}{|E_r|}\right)^{b_J}\lim_{r\to\infty}\binom{k_{r,m}}{b_1,b_2,\ldots,b_J}$$

$$\overset{(c)}{=}\lim_{r\to\infty}\prod_{j=1}^{J}(q_{j,r})^{b_j}\lim_{r\to\infty}\binom{k_{r,m}}{b_1,b_2,\ldots,b_J}$$

$$=\prod_{j=1}^{J}(q_j)^{b_j}\lim_{r\to\infty}\binom{k_{r,m}}{b_1,b_2,\ldots,b_J}$$

where equality (a) follows since for all $j\in\{1,\ldots,J\}$, $|E_r^{[j]}|\to\infty$ as we let r tend to infinity, while on the other hand, the maximal right degree (and hence, also $b_1,\ldots,b_J$

and $k_{r,m}$) stay bounded; equality (b) is valid due to the constraint $\sum_{j=1}^{J} b_j = k_{r,m}$, and equality ($c$) follows from (3.21). ■

By letting r tend to infinity on both sides of (3.30), and substituting (3.31) and (3.32) in the RHS of (3.30), we get that for all $p \in \mathbb{N}$

$$\begin{aligned}
&\lim_{r\to\infty} \mathbb{E}\left[\prod_{j=1}^{J} (g_{j,p})^{\beta_{r,j,m}}\right] \\
&\overset{(a)}{=} \mathbb{E}\left[\lim_{r\to\infty} \mathbb{E}\left(\prod_{j=1}^{J} (g_{j,p})^{\beta_{r,j,m}} \,\Big|\, \sum_{j=1}^{J} \beta_{r,j,m} = k_{r,m}\right)\right] \\
&\overset{(b)}{=} \mathbb{E}\left[\lim_{r\to\infty} \sum_{\substack{b_1,\ldots,b_J \geq 0 \\ \sum_{j=1}^{J} b_j = k_{r,m}}} \binom{k_{r,m}}{b_1, b_2, \ldots, b_J} \prod_{j=1}^{J} (q_j\, g_{j,p})^{b_j}\right] \\
&= \mathbb{E}\left[\lim_{r\to\infty} \left(\sum_{j=1}^{J} q_j g_{j,p}\right)^{k_{r,m}}\right] \\
&\overset{(c)}{=} \sum_{k=1}^{d_{\mathrm{c,max}}} \left\{\Gamma_k \left(\sum_{j=1}^{J} q_j g_{j,p}\right)^{k}\right\} \\
&= \Gamma\left(\sum_{j=1}^{J} q_j g_{j,p}\right)
\end{aligned} \tag{3.33}$$

where equality (a) follows from (3.30) and since the right degree distribution is independent of r (note that the outer expectation in equality (a) is performed w.r.t. the degree of the m^{th} parity-check node); equality (b) follows from (3.31) and (3.32), and since the number of terms in the sum is bounded (this number is upper bounded by $(k_{r,m})^{J-1}$, so it is bounded for all $r \in \mathbb{N}$ due to the fact that the maximal right degree is fixed), and equality (c) follows since the right degree distribution is independent of r. Since the limit in (3.33) does not depend on the index m which appears in the inner summation at the LHS of (3.29) and also $\lim_{r\to\infty} n_r(1 - R_{\mathrm{d}}) = \infty$, then we get from (3.33)

$$\begin{aligned}
&\lim_{r\to\infty} \frac{1}{n_r(1-R_{\mathrm{d}})} \sum_{m=1}^{n_r(1-R_{\mathrm{d}})} \mathbb{E}\left[\prod_{j=1}^{J} (g_{j,p})^{\beta_{r,j,m}}\right] \\
&\overset{(a)}{=} \lim_{r\to\infty} \mathbb{E}\left[\prod_{j=1}^{J} (g_{j,p})^{\beta_{r,j,m}}\right] \\
&\overset{(b)}{=} \Gamma\left(\sum_{j=1}^{J} q_j g_{j,p}\right)
\end{aligned} \tag{3.34}$$

where equality (a) follows from the fact that if $\{a_r\}$ is a convergent sequence then the equality $\lim_{r\to\infty} \frac{1}{r}\sum_{i=1}^{r} a_i = \lim_{r\to\infty} a_r$ holds, and also since any sub-sequence of a convergent sequence converges to the same limit as of the original sequence; equality (b) follows from (3.33). Combining (3.29) and (3.34) gives

$$1 - \sum_{j=1}^{J} p_j C_j - (1 - R_{\mathrm{d}}) \left(1 - \frac{1}{2\ln 2} \sum_{p=1}^{\infty} \left\{ \frac{1}{p(2p-1)} \, \Gamma\left(\sum_{j=1}^{J} q_j g_{j,p} \right) \right\} \right) \leq 0.$$

Finally, solving the last inequality for R_{d} gives the upper bound on the design rate in (3.27). ■

Example 3.1 For the special case where the J parallel MBIOS channels are binary erasure channels where the erasure probability of the j^{th} channel is ε_j, we get from (3.3)

$$g_{j,p} = 1 - \varepsilon_j \,, \quad \forall\, j \in \{1, \ldots, J\},\ p \in \mathbb{N}. \tag{3.35}$$

Since $g_{j,p}$ is independent of p for a BEC, and based on the equality $\sum_{p=1}^{\infty} \frac{1}{2p(2p-1)} = \ln 2$, we obtain from Theorem 3.1 that the common design rate of the sequence of LDPC ensembles is upper bounded by

$$R_{\mathrm{d}} \leq 1 - \frac{\displaystyle\sum_{j=1}^{J} p_j \varepsilon_j}{1 - \Gamma\left(1 - \displaystyle\sum_{j=1}^{J} q_j \, \varepsilon_j \right)}. \tag{3.36}$$

This result coincides with [69, Theorem 2].

The proof of Theorem 3.1 relies on the assumption that the right degree distribution ρ is fixed, and does not depend on the ordinal number r of the ensemble. For a capacity-achieving sequence of LDPC ensembles, both the maximal and the average right degrees tend to infinity (see [81, Theorem 1]). Hence, for a capacity-achieving sequence of LDPC codes, ρ cannot be fixed.

Remark 3.2 One can think of Lemma 3.5 in terms of drawing colored balls from an urn (where the colors are determined in one to one correspondence with the assignments of the various edges to the component channels). Since an edge can only be assigned once to a parity-check node, the balls are not returned to the urn after they are chosen. As the block length tends to infinity, so does the number of edges originating in each of the parallel channels (this is the reason for requiring that q_j is positive for all j). Since the degree of the parity-check nodes remains finite, we are drawing a finite number of balls from an urn which contains an infinite number

of balls of each color. This lemma simply says that drawing without replacement is equivalent to drawing with replacement if the number of draws is finite and the number of balls of each color becomes infinite. Note that this result looks rather intuitive from a statistical point of view.

Remark 3.3 We wish to discuss a possible refinement of the statement in Theorem 3.1. Let us assume that the (overall) degree distributions λ and ρ are fixed, but due to the transmission over parallel channels, the corresponding vector of degree distributions $\lambda_{\mathbf{r}} = (\lambda_r^{[1]}, \ldots, \lambda_r^{[J]})$ and also $p_{j,r}$ and $q_{j,r}$ depend on the code from the ensemble (n_r, λ, ρ). Since the derivation of this theorem relies on the bounds on the conditional entropy from Section 3.2 (which are valid code by code), one can refine the statement in Theorem 3.1 so that the modified theorem permits the dependency of the vector $(\lambda_r^{[1]}, \ldots, \lambda_r^{[J]})$ on the specific code chosen from the ensemble. In this case, the equalities in (3.26) are transformed to

$$p_j = \lim_{r\to\infty} \mathbb{E}\left[p_{j,r}(\mathcal{C})\right], \quad q_j = \lim_{r\to\infty} \mathbb{E}\left[q_{j,r}(\mathcal{C})\right]$$

where the expectation is carried over the code $\mathcal{C}$ from the ensemble (n_r, λ, ρ). In this case, the proof of Theorem 3.1 involves an expectation over $\mathcal{C}$ on both sides of (3.28) (which is valid code by code) and then we let r tend to infinity, as in (3.29). By invoking Jensen's inequality, Lemma 3.5 is changed under the above assumption to the inequality

$$\lim_{r\to\infty} \mathbb{E}_{\mathcal{C}} \left[\frac{\binom{|E_r^{[1]}|}{b_1} \cdots \binom{|E_r^{[J]}|}{b_J}}{\binom{|E_r|}{k_{r,m}}} \right] \geq \prod_{j=1}^{J} (q_j)^{b_j} \lim_{r\to\infty} \binom{k_{r,m}}{b_1, b_2, \ldots, b_J}$$

and correspondingly, (3.33) is changed to

$$\lim_{r\to\infty} \mathbb{E}_{\mathcal{C}} \left[\prod_{j=1}^{J} (g_{j,p})^{\beta_{r,j,m}} \right] \geq \Gamma\left(\sum_{j=1}^{J} q_j g_{j,p} \right).$$

Therefore, the upper bound on the design rate in (3.27) holds for the more general setting as above.

3.4 Achievable Rates of Punctured LDPC Codes

In this section we derive upper bounds on the achievable rates of punctured LDPC codes whose transmission takes place over an MBIOS channel, and the codes are ML decoded. The analysis in this section relies on the bound presented in Section 3.3.

Let $\mathcal{C}$ be a binary linear block code. Assume its code bits are partitioned into J disjoint sets, and the bits of the j^{th} set are randomly punctured with a puncturing rate π_j (where $j \in \{1, \ldots, J\}$). The transmission of this code over an MBIOS channel is equivalent to transmitting the code over a set of J parallel MBIOS channels where each of these channels forms a serial concatenation of a BEC whose erasure probability is equal to the puncturing rate π_j, followed by the original MBIOS channel (see e.g., [35, 65, 68, 69]).

3.4.1 Some Preparatory Lemmas

This sub-section presents two lemmas which are later used to prove results for ensembles of randomly and intentionally punctured LDPC codes (denoted by RP-LDPC and IP-LDPC codes, respectively).

In the following lemma, we consider a punctured linear block code and provide an upper bound on the conditional entropy of a codeword before puncturing, given the received sequence at the output of the channel. This upper bound is expressed in terms of the bit error probability of the punctured code.

Lemma 3.6 Let $\mathcal{C}'$ be a binary linear block code of length n and rate R', and let $\mathcal{C}$ be a code which is obtained from $\mathcal{C}'$ by puncturing some of its code bits. Assume that the transmission of the code $\mathcal{C}$ takes place over an arbitrary communication channel, and the code is decoded by an arbitrary decoding algorithm. Let $\mathbf{X}' = (X'_1, \ldots, X'_n)$ and $\mathbf{Y} = (Y_1, \ldots, Y_n)$ (where the punctured bits are replaced by question marks which have an LLR of zero) designate the transmitted codeword of $\mathcal{C}'$ and the received sequence, respectively. Then, the conditional entropy of the original codeword of $\mathcal{C}'$ given the received sequence satisfies

$$\frac{H(\mathbf{X}'|\mathbf{Y})}{n} \leq R' \, h_2(P_{\text{b}}) \tag{3.37}$$

where P_{b} designates the bit error probability of the punctured code $\mathcal{C}$.

Proof: The proof follows directly from Lemma 3.2, and the equivalence between the transmission of punctured codes over an MBIOS channel and the special case of transmitting these codes over a set of parallel channels (see the introductory paragraph of Section 3.4). ■

Puncturing serves to increase the rate of the original code by reducing the length of the codeword. It may however cause several codewords to be mapped onto a single codeword, thereby reducing the dimension of the code. Consider a binary linear code, $\mathcal{C}'$, of length n and rate R' and assume a fraction γ of its code bits are punctured. In

the case that the dimension is not reduced by puncturing, the rate of the punctured code is given by $R = \frac{R'}{1-\gamma}$. In the general case, we cannot guarantee that the dimension of the code is not reduced. However, for a sequence of punctured codes whose bit error probability vanishes as the block length of the codes tends to infinity, the following lemma shows that the rate of the punctured codes converges to the desired rate R.

Lemma 3.7 Let $\{\mathcal{C}'_r\}$ be a sequence of binary linear block codes of length n_r and rate R'_r, and let $\{\mathcal{C}_r\}$ be a sequence of codes which is obtained from $\{\mathcal{C}'_r\}$ by puncturing a fraction γ of the code bits. Assume the sequence of punctured codes $\{\mathcal{C}_r\}$ achieves vanishing bit error probability in the limit where we let r tend to infinity. Then, the asymptotic rate R of the sequence of punctured codes is given by

$$R = \frac{R'}{1-\gamma} \tag{3.38}$$

where $R' = \lim_{r\to\infty} R'_r$ is the asymptotic rate of the original sequence of codes $\{\mathcal{C}'_r\}$.

Proof: Let $\mathbf{X'_r} = (X'_1, \ldots, X'_{n_r})$ and $\mathbf{Y_r} = (Y_r \ldots, Y_{n_r})$ designate the original codeword (before puncturing) and the received sequence (after puncturing), respectively. Since we assume the there exists a decoding algorithm such that the punctured codes achieve vanishing bit error probability, we have from lemma 3.6 that

$$\lim_{r\to\infty} \frac{H(\mathbf{X'_r}|\mathbf{Y_r})}{n_r} = 0.$$

Let $\mathbf{X_r} = (X_1, \ldots, X_{n_r})$ designate the codeword after puncturing (where the punctured bits are replaced by question marks). Since $\mathbf{X'_r} \Rightarrow \mathbf{X_r} \Rightarrow \mathbf{Y_r}$ forms a Markov chain, then by the information processing inequality, we get $H(\mathbf{X'_r}|\mathbf{X_r}) \leq H(\mathbf{X'_r}|\mathbf{Y_r})$. The non-negativity of the conditional entropy therefore yields that

$$\lim_{r\to\infty} \frac{H(\mathbf{X'_r}|\mathbf{X_r})}{n_r} = 0. \tag{3.39}$$

Denote the number of dimensions of the codes $\mathcal{C}'_r$ and $\mathcal{C}_r$ by d'_r and d_r, respectively. Since $\mathcal{C}'_r$ is binary and linear, every codeword of $\mathcal{C}_r$ originates from exactly $2^{d'_r - d_r}$ different codewords of $\mathcal{C}'_r$. The codewords are assumed to be transmitted with equal probability, and therefore $H(\mathbf{X'_r}|\mathbf{X_r}) = d'_r - d_r$. Let R_r designate the rate of the punctured code $\mathcal{C}_r$. By definition, $d'_r = R'_r n_r$, and since $n_r(1-\gamma)$ forms the block length of the punctured code $\mathcal{C}_r$, then $d_r = R_r n_r (1-\gamma)$. Substituting the last three equalities into (3.39) gives

$$\lim_{r\to\infty} (R'_r - R_r(1-\gamma)) = 0.$$

This completes the proof of the lemma. ■

For a sequence of codes $\{\mathcal{C}'_r\}$, it is natural to refer to their code rates R'_r. However, for sequences of ensembles, where parity-check matrices are randomly picked, such matrices are unlikely to be full rank. Hence, a more natural approach is to refer to their design rates. To this end, we define the design rate of codes which are obtained by puncturing some code bits of binary linear block codes.

Definition 3.1 Let $\mathcal{C}'$ be a binary linear block code of length n, H' be a $c \times n$ parity-check matrix of $\mathcal{C}'$ and $R'_d \triangleq 1 - \frac{c}{n}$ designate the design rate of the code $\mathcal{C}'$. Let $\mathcal{C}$ be a code which is obtained from $\mathcal{C}'$ by puncturing a fraction γ of the code bits. The *design rate* of $\mathcal{C}$ is defined as

$$R_{\mathrm{d}} \triangleq \frac{R'_{\mathrm{d}}}{1-\gamma}. \tag{3.40}$$

From Lemma 3.7, it follows that for an arbitrary sequence of punctured codes which achieves vanishing bit error probability, their asymptotic design rate is equal in probability 1 to their asymptotic rate if and only if this condition also holds for the original sequence of codes before their puncturing. For un-punctured ensembles of LDPC codes, a sufficient condition for the asymptotic convergence of the rate to the design rate is introduced in [60, Lemma 7] (which is also presented in the preliminaries of our companion paper as [118, Lemma 2.1]). In Section 3.4.4, we apply this lemma to show that the bounds on the achievable rates of ensembles of punctured LDPC codes apply to their actual code rates and not only to their asymptotic design rates.

3.4.2 Randomly Punctured LDPC Codes

In this section, we consider the achievable rates of randomly punctured LDPC (RP-LDPC) codes. We assume that the transmission of these codes takes place over an MBIOS channel, and refer to their achievable rates under optimal ML decoding. The upper bound on the achievable rates of ensembles of RP-LDPC codes relies on the analysis in Section 3.3 where we derived an upper bound on the achievable rates of LDPC codes for parallel channels.

In the following, we assume that the communication takes place over an MBIOS channel with capacity C, and define

$$g_p \triangleq \int_0^\infty a(l)\,(1+e^{-l})\tanh^{2p}\left(\frac{l}{2}\right)\mathrm{d}l\,, \qquad p \in \mathbb{N} \tag{3.41}$$

where a designates the pdf of the LLR of the channel given that its input is zero.

Theorem 3.2 Let $\{(n_r, \lambda, \rho)\}_{r=1}^{\infty}$ be a sequence of ensembles of binary LDPC codes whose block length (n_r) tends to infinity as $r \to \infty$. Assume that a sequence of

ensembles of RP-LDPC codes is constructed in the following way: for each code from an ensemble of the original sequence, a subset of αn_r code bits is a-priori selected, and these bits are randomly punctured at a fixed rate (P_{pct}). Assume that the punctured codes are transmitted over an MBIOS channel with capacity C, and that in the limit where r approaches infinity, the sequence of ensembles of RP-LDPC codes achieves vanishing bit error probability under some decoding algorithm. Then in probability 1 w.r.t. the random puncturing patterns, the asymptotic design rate (R_{d}) of the new sequence satisfies

$$R_{\text{d}} \leq \frac{1}{1-\alpha P_{\text{pct}}} \left(1 - \frac{1-(1-\alpha P_{\text{pct}})C}{1 - \frac{1}{2\ln 2} \sum_{p=1}^{\infty} \left\{ \frac{1}{p(2p-1)} \Gamma\Big((1-P_{\text{pct}}+\xi) g_p \Big) \right\}} \right) \tag{3.42}$$

where Γ denotes the right degree distribution (from the node perspective) of the original sequence, g_p is introduced in (3.41), and ξ is the following positive number:

$$\xi \triangleq 2(1-\alpha) P_{\text{pct}} \int_0^1 \lambda(x)\, \mathrm{d}x. \tag{3.43}$$

Proof: By assumption, we select a set of code bits whose size is a fraction α of the n_r code bits, and these bits are randomly punctured at rate P_{pct}. The transmission of the resulting codeword over an MBIOS channel is equivalent to the transmission of the original codeword over a set of $J = 2$ parallel channels. The first channel, referring to the set of code bits which are randomly punctured, is a serial concatenation of a BEC with erasure probability P_{pct} and the original MBIOS channel; the second channel which refers to the rest of the bits (which are transmitted without being randomly punctured) is the original MBIOS channel. For simplicity, let us first assume that the degree distribution associated with the selected subset of αn_r code bits which are randomly punctured is independent of the specific code from the ensemble (n_r, λ, ρ). Based on the discussion above and the notation in Section 3.3, the transmission of the n_r code bits over these two parallel channels induces a sequence of ensembles of LDPC codes, $\{(n_r, \lambda_{\mathbf{r}}, \rho)\}_{r=1}^{\infty}$, where $\lambda_{\mathbf{r}} = (\lambda_r^{[1]}, \lambda_r^{[2]})$ depends on the selection of the subset of αn_r code bits which are randomly punctured. Following this equivalence, we get from the notation in Theorem 3.1 that

$$\begin{aligned} & p_1 = \alpha, \quad p_2 = 1-\alpha, \quad C_1 = C(1-P_{\text{pct}}), \quad C_2 = C \\ \Rightarrow \quad & \sum_{j=1}^{J} p_j C_j = C(1-\alpha P_{\text{pct}}). \end{aligned} \tag{3.44}$$

In order to apply Theorem 3.1 to our case, we find a global lower bound on the sum $\sum_{j=1}^{J} q_j g_{j,p}$ which does not depend on the a-priori selection of the subset of code bits which are randomly punctured. From (3.3) and (3.41), it follows that for all $p \in \mathbb{N}$:

$$\begin{aligned} g_{1,p} &= \int_0^\infty \left[P_{\text{pct}}\delta(l) + (1-P_{\text{pct}})a(l)\right](1+e^{-l})\tanh^{2p}\left(\frac{l}{2}\right)\,\mathrm{d}l \\ &= (1-P_{\text{pct}})\int_0^\infty a(l)(1+e^{-l})\tanh^{2p}\left(\frac{l}{2}\right)\,\mathrm{d}l \\ &= (1-P_{\text{pct}})\,g_p \end{aligned}$$

and $g_{2,p} = g_p$. Based on Lemmas 3.3 and 3.4, we get that for all $p \in \mathbb{N}$

$$q_1 g_{1,p} + q_2 g_{2,p} = \frac{\alpha g_p (1-P_{\text{pct}})\displaystyle\int_0^1 \lambda(x)\,\mathrm{d}x}{\displaystyle\int_0^1 \lambda_r^{[1]}(x)\,\mathrm{d}x} + \frac{(1-\alpha)g_p\displaystyle\int_0^1 \lambda(x)dx}{\displaystyle\int_0^1 \lambda_r^{[2]}(x)\,\mathrm{d}x} \tag{3.45}$$

where the following constraint is satisfied:

$$\frac{\alpha}{\displaystyle\int_0^1 \lambda_r^{[1]}(x)\,\mathrm{d}x} + \frac{1-\alpha}{\displaystyle\int_0^1 \lambda_r^{[2]}(x)\,\mathrm{d}x} = \frac{1}{\displaystyle\int_0^1 \lambda(x)\,\mathrm{d}x} \tag{3.46}$$

and

$$\int_0^1 \lambda_r^{[1]}(x)\,\mathrm{d}x \le \frac{1}{2}\,, \quad \int_0^1 \lambda_r^{[2]}(x)\,\mathrm{d}x \le \frac{1}{2} \tag{3.47}$$

due to the fact that $\lambda^{[1]}(x) \le x$ and $\lambda^{[2]}(x) \le x$ for $x \in [0,1]$ (even without explicitly knowing $\lambda^{[1]}$ and $\lambda^{[2]}$ which depend on the a-priori choice of the subset of bits which are randomly punctured). Based on (3.45)–(3.47), we get

$$\begin{aligned} &q_1 g_{1,p} + q_2 g_{2,p} \\ &= (1-P_{\text{pct}})g_p \int_0^1 \lambda(x)\,\mathrm{d}x \left(\frac{\alpha}{\displaystyle\int_0^1 \lambda_r^{[1]}(x)\,\mathrm{d}x} + \frac{1-\alpha}{\displaystyle\int_0^1 \lambda_r^{[2]}(x)\,\mathrm{d}x}\right) \\ &\qquad + \frac{(1-\alpha)P_{\text{pct}}g_p\displaystyle\int_0^1 \lambda(x)\,\mathrm{d}x}{\displaystyle\int_0^1 \lambda_r^{[2]}(x)\,\mathrm{d}x} \\ &= (1-P_{\text{pct}})g_p + \frac{(1-\alpha)P_{\text{pct}}g_p\displaystyle\int_0^1 \lambda(x)\,\mathrm{d}x}{\displaystyle\int_0^1 \lambda_r^{[2]}(x)\,\mathrm{d}x} \\ &\ge \left(1 - P_{\text{pct}} + 2(1-\alpha)P_{\text{pct}}\int_0^1 \lambda(x)\,\mathrm{d}x\right) g_p \\ &= (1 - P_{\text{pct}} + \xi)g_p \end{aligned}$$

where ξ is defined in (3.43). Since the degree distribution Γ is a monotonic increasing function, then

$$\Gamma\left(\sum_{j=1}^{J} q_j g_{j,p}\right) \geq \Gamma\big((1 - P_{\text{pct}} + \xi) g_p\big). \tag{3.48}$$

By substituting (3.44) and (3.48) in the RHS of (3.27), we obtain the following upper bound on the asymptotic design rate of the original sequence

$$R'_{\text{d}} \leq 1 - \frac{1 - (1 - \alpha P_{\text{pct}})C}{1 - \dfrac{1}{2 \ln 2} \displaystyle\sum_{p=1}^{\infty} \left\{ \frac{1}{p(2p-1)} \Gamma\Big((1 - P_{\text{pct}} + \xi) g_p\Big) \right\}}.$$

Since as $r \to \infty$, in probability 1 w.r.t. the puncturing patterns, a fraction $\gamma = \alpha P_{\text{pct}}$ of the code bits are punctured, then the asymptotic design rate (R_{d}) of this sequence satisfies the equality

$$R_{\text{d}} = \frac{R'_{\text{d}}}{1 - \alpha P_{\text{pct}}} \tag{3.49}$$

from which the theorem follows.

For the case where the degree distribution associated with the subset of code bits which are randomly punctured depends on the code $\mathcal{C}$ from the ensemble (n_r, λ, ρ), the pair $(\lambda_r^{[1]}, \lambda_r^{[2]})$ cannot be considered to be uniform over all the codes from this ensemble. In this case, Theorem 3.1 is not directly applicable. In order to circumvent the problem, we rely on the discussion in Remark 3.3, and on the fact that the lower bound on $q_1 g_{1,p} + q_2 g_{2,p}$ which is given above in terms of ξ from (3.43) is universal for all the codes from this ensemble (i.e., it only depends on λ, but does not depend on the specific degree distributions $\lambda_r^{[1]}(\mathcal{C})$ and $\lambda_r^{[2]}(\mathcal{C})$ which are associated with the code $\mathcal{C}$ from the ensemble). In light of this reasoning, the proof of the theorem for ensembles of RP-LDPC codes also follows in the more general setting where the degree distribution associated with the subset of the code bits which are randomly punctured depends on the specific code from the ensemble. ■

Remark 3.4 Note that in the above proof, we derive an upper bound on the number of edges adjacent to variable nodes which are punctured in probability P_{pct}; this is done by assuming that the degree of all the un-punctured nodes is 2 (which is the minimal possible degree for a variable node), and counting the number of the remaining edges. In the case that the original codes before puncturing have a minimal variable degree of $\Lambda_{\min} > 2$, the upper bound can be tightened by assuming that each un-punctured node is of degree $\Lambda_{\min}$. This results in replacing ξ in (3.43) with $\xi' \triangleq \Lambda_{\min}(1 - \alpha) P_{\text{pct}} \int_0^1 \lambda(x)\, \mathrm{d}x$.

3.4.3 Intentionally Punctured LDPC Codes

In [35], Ha et al. show that good codes can be constructed by puncturing good ensembles of LDPC codes using a technique called "intentional puncturing". In this approach, the code bits are partitioned into disjoint sets so that each set contains all the code bits whose corresponding variable nodes have the same degree. The code bits in each of these sets are randomly punctured at a fixed puncturing rate.

We briefly present the notation used in [35] for the characterization of ensembles of intentionally punctured LDPC (IP-LDPC) codes. Consider an ensemble of LDPC codes with left and right edge degree distributions λ and ρ, respectively. For each degree j such that $\lambda_j > 0$, a puncturing rate $\pi_j \in [0,1]$ is determined for randomly puncturing the set of code bits which correspond to variable nodes of degree j. The polynomial associated with this puncturing pattern is

$$\pi^{(0)}(x) \triangleq \sum_{j=1}^{\infty} \pi_j x^{j-1}. \tag{3.50}$$

An ensemble of IP-LDPC codes can be therefore represented by the quadruplet $(n, \lambda, \rho, \pi^{(0)})$ where n designates the block length of these codes, λ and ρ are the left and right degree distributions from the edge perspective, respectively, and $\pi^{(0)}$ is the polynomial which corresponds to the puncturing pattern, as given in (3.50). The average fraction of punctured bits is given by $p^{(0)} = \sum_{j=1}^{\infty} \Lambda_j \pi_j$ where Λ is the left node degree distribution of the original LDPC ensemble. The following statement, which relies on Theorem 3.1, provides an upper bound on the common design rate of a sequence of ensembles of IP-LDPC codes. This bound refers to ML decoding (and hence, to any sub-optimal decoding algorithm).

Theorem 3.3 Let $\{(n_r, \lambda, \rho, \pi^{(0)})\}_{r=1}^{\infty}$ be a sequence of ensembles of IP-LDPC codes transmitted over an MBIOS channel, and assume that n_r tends to infinity as $r \to \infty$. Let C be the channel capacity, and a be the pdf of the LLR at the output of the channel given its input is zero. If the asymptotic bit error probability of this sequence vanishes under ML decoding (or any sub-optimal decoding algorithm) as $r \to \infty$, then in probability 1 w.r.t. the puncturing patterns, the asymptotic design rate R_{d} of these ensembles satisfies

$$R_{\mathrm{d}} \leq \frac{1}{1-p^{(0)}} \left[1 - \frac{1-(1-p^{(0)})C}{1 - \frac{1}{2\ln 2} \sum_{p=1}^{\infty} \left\{ \frac{1}{p(2p-1)} \Gamma\left(\left(1 - \sum_{j=1}^{\infty} \lambda_j \pi_j \right) g_p \right) \right\}} \right] \tag{3.51}$$

where Γ denotes the right degree distribution from the node perspective,

$$p^{(0)} \triangleq \sum_{j=1}^{\infty} \Lambda_j \pi_j \tag{3.52}$$

designates the average puncturing rate of the code bits, and g_p is the functional of the MBIOS channel introduced in (3.41).

Proof: The proof follows from Theorem 3.1, and the observation that IP-LDPC codes form a special case of the ensemble (n, λ, ρ) examined in Section 3.3. For a sequence of ensembles of IP-LDPC codes, $\{(n_r, \lambda, \rho, \pi^{(0)})\}$, the number of parallel MBIOS channels used for transmission is equal to the number of strictly positive coefficients in the polynomial λ, i.e., $J \triangleq |\{i : \lambda_i > 0\}|$. Denote these degrees by $i_1, \ldots, i_J$, then the bits transmitted over the j^{th} channel are those involved in exactly i_j parity-check equations (i.e., the bits whose corresponding variable nodes are of degree i_j). From the above discussion, it follows that the fraction of code bits transmitted over the j^{th} channel is given by

$$p_j = \Lambda_{i_j}, \quad j \in \{1, \ldots, J\} \tag{3.53}$$

and the fraction of edges in the bipartite graph which are connected to variable nodes transmitted of the j^{th} channel is given by

$$q_j = \lambda_{i_j}, \quad j \in \{1, \ldots, J\}. \tag{3.54}$$

The transmission of IP-LDPC codes over an MBIOS channel is equivalent to transmitting the original codes (before puncturing) over a set of J parallel MBIOS channels where each of these channels is formed by a serial concatenation of a BEC whose erasure probability is equal to the puncturing rate π_{i_j}, followed by the original MBIOS channel. Hence, the pdf of the LLR at the output of the j^{th} MBIOS channel given its input is 1 gets the form

$$a(l; j) = \pi_{i_j} \delta_0(l) + (1 - \pi_{i_j}) a(l), \quad l \in \mathbb{R} \tag{3.55}$$

and the capacity of this channel is

$$C_j = C(1 - \pi_{i_j}). \tag{3.56}$$

By substituting (3.55) into (3.3), we get that for all $j \in \{1, \ldots, J\}$ and $p \in \mathbb{N}$

$$\begin{aligned} g_{j,p} &= \int_0^{\infty} \left[\pi_{i_j} \delta_0(l) + (1 - \pi_{i_j}) a(l)\right] (1 + e^{-l}) \tanh^{2p}\left(\frac{l}{2}\right) \mathrm{d}l \\ &= (1 - \pi_{i_j}) \int_0^{\infty} a(l) (1 + e^{-l}) \tanh^{2p}\left(\frac{l}{2}\right) \mathrm{d}l \\ &= (1 - \pi_{i_j}) g_p \end{aligned} \tag{3.57}$$

where the last equality is based on (3.41). The statement now follows by substituting (3.53), (3.54), (3.56) and (3.57) in (3.27); we use the scaling factor for the design rate of the punctured codes, as given in Definition 3.1. In this case, the parameter γ tends to the average puncturing rate $p^{(0)}$ of the code bits, as defined in (3.52), where this convergence is in probability 1 w.r.t. the puncturing patterns. Finally, since $\lambda_j = \Lambda_j = 0$ for $j \notin \{i_1, \ldots, i_J\}$, then regarding the sums in the RHS of (3.51), we get the equalities $\sum_{j=1}^{\infty} \Lambda_j \pi_j = \sum_{j=1}^{J} \Lambda_{i_j} \pi_{i_j}$ and $\sum_{j=1}^{\infty} \lambda_j \pi_j = \sum_{j=1}^{J} \lambda_{i_j} \pi_{i_j}$. This completes the proof of the theorem. ■

Remark 3.5 Let us consider a more general case of punctured ensembles of LDPC codes where the original code bits are split into J arbitrary sets and each set is punctured at a different rate. For this general case, it is possible to apply Theorem 3.1 to derive an upper bound on the achievable rates which only depends on the expected fractions of punctured code bits and edges in the graph attached to variable nodes of punctured bits. Theorems 3.2 and 3.3 emerge as corollaries of such a theorem (in this paper we do not take this approach since we analyze two strategies of puncturing as special cases of transmission over parallel channels). In the case of ensembles of RP-LDPC codes, the fraction of edges adjacent to punctured bits is not known in general. Hence, for the derivation of upper bounds on their achievable rates, we employ a lower bound on the fraction of edges adjacent to punctured bits in a similar way to the proof of Theorem 3.2.

3.4.4 Numerical Results for Intentionally Punctured LDPC Codes

In this section, we present a comparison between thresholds under iterative message-passing decoding and bounds on thresholds under ML decoding for ensembles of IP-LDPC codes. It is assumed that the transmission of the punctured LDPC codes takes place over a binary-input AWGN channel. The pairs of degree distributions and the corresponding puncturing patterns were originally presented in [34, 35]. We use these ensembles in order to study their inherent gap to capacity, and also study how close to optimal iterative decoding is for these ensembles (in the asymptotic case where the block length goes to infinity).

We refer here to three ensembles of IP-LDPC codes: Tables 3.1 and 3.2 refer to two ensembles of rate$-\frac{1}{2}$ LDPC codes which by puncturing, their rates vary between 0.50 and 0.91; Table 3.3 refers to an ensemble of rate$-\frac{1}{10}$ LDPC codes which by puncturing, its rate varies between 0.10 and 0.83. Based on [60, Lemma 7], we verify

$\pi^{(0)}(x)$ (puncturing pattern)	Design rate	Capacity limit	Lower bound (ML decoding)	Iterative (IT) Decoding	Fractional gap to capacity (ML vs. IT)
0	0.500	0.187 dB	0.270 dB	0.393 dB	$\geq 40.3\%$
$0.07886x+0.01405x^2+0.06081x^3+0.07206x^9$	0.528	0.318 dB	0.397 dB	0.526 dB	$\geq 37.9\%$
$0.20276x+0.09305x^2+0.03356x^3+0.16504x^9$	0.592	0.635 dB	0.716 dB	0.857 dB	$\geq 36.4\%$
$0.25381x+0.15000x^2+0.34406x^3+0.019149x^9$	0.629	0.836 dB	0.923 dB	1.068 dB	$\geq 37.3\%$
$0.31767x+0.18079x^2+0.05265x^3+0.24692x^9$	0.671	1.083 dB	1.171 dB	1.330 dB	$\geq 35.6\%$
$0.36624x+0.24119x^2+0.49649x^3+0.27318x^9$	0.719	1.398 dB	1.496 dB	1.664 dB	$\geq 36.9\%$
$0.41838x+0.29462x^2+0.05265x^3+0.30975x^9$	0.774	1.814 dB	1.927 dB	2.115 dB	$\geq 37.2\%$
$0.47074x+0.34447x^2+0.02227x^3+0.34997x^9$	0.838	2.409 dB	2.547 dB	2.781 dB	$\geq 37.1\%$
$0.52325x+0.39074x^2+0.01324x^3+0.39436x^9$	0.912	3.399 dB	3.607 dB	3.992 dB	$\geq 35.1\%$

Table 3.1: Comparison of thresholds for ensembles of IP-LDPC codes where the original ensemble before puncturing has the degree distributions $\lambda(x) = 0.25105x + 0.30938x^2 + 0.00104x^3 + 0.43853x^9$ and $\rho(x) = 0.63676x^6 + 0.36324x^7$ (so its design rate is equal to $\frac{1}{2}$). The transmission of these codes takes place over a binary-input AWGN channel. The table compares values of $\frac{E_b}{N_0}$ referring to the capacity limit, the bound given in Theorem 3.3 (which provides a lower bound on $\frac{E_b}{N_0}$ under ML decoding), and thresholds under message-passing decoding. The fractional gap to capacity (see the rightmost column) measures the ratio of the gap to capacity under optimal ML decoding and the achievable gap to capacity under (sub-optimal) message-passing decoding. The pair of degree distributions for the ensemble of LDPC codes, and the polynomials which correspond to its puncturing patterns are taken from [35, Table 2].

that the design rates of these three ensembles of LDPC codes (before puncturing) are equal in probability 1 to the asymptotic rates of codes from these ensembles. This conclusion still holds for the punctured LDPC ensembles given in Tables 3.1–3.3 (see Lemma 3.7). This enables to calculate the capacity limits which refer to the design rates of these ensembles, and to evaluate the gaps to capacity under ML decoding and iterative decoding for these ensembles of punctured LDPC codes.

For various ensembles of IP-LDPC codes, Tables 3.1–3.3 provide lower bounds on the inherent gap to capacity under optimal ML decoding (based on Theorem 3.3); these values are compared to the corresponding gaps to capacity under message-passing decoding (whose calculation is based on the density evolution analysis). On one hand, Tables 3.1–3.3 provide a quantitative assessment of the loss in the asymptotic performance which is attributed to the sub-optimality of iterative decoding (as compared to ML decoding), and on the other hand, they provide an assessment of the inherent loss in performance which is attributed to the structure of the ensembles, even if ML decoding could be applied to decode these codes. The loss in performance in both cases is measured in terms of $\frac{E_{\mathrm{b}}}{N_0}$ in decibels. It is demonstrated in Tables 3.1–3.3 that for various good ensembles of IP-LDPC codes, the asymptotic loss in performance due to the code structure is still non-negligible as compared to the corresponding loss due to the sub-optimality of iterative decoding. As an example, for all the ensembles of IP-LDPC codes considered in Table 3.1 (which were originally introduced in [35, Table 2]), the gap to capacity under the sum-product iterative decoding algorithm does not exceed 0.6 dB; however, under ML decoding, the gap to capacity is always greater than $\frac{1}{3}$ of the corresponding gap to capacity under this iterative decoding algorithm; therefore, the results in Table 3.1 regarding the thresholds under ML decoding further emphasize the efficiency of the sum-product decoding algorithm for these ensembles, especially in light of its moderate complexity.

Tables 3.1–3.3 also show that the performance of the punctured LDPC codes is degraded at high rates, where one needs to pay a considerable penalty for using punctured codes. This phenomenon was explained in [68, Theorem 1] by the threshold effect for ensembles of punctured LDPC codes.

Following the performance analysis of punctured LDPC codes in [34, 35, 39, 68], the numerical results shown in Tables 3.1–3.3 exemplify the high potential of puncturing in designing codes which operate closely to the Shannon capacity limit and are used for rate-compatible coding for various MBIOS channels. Other examples of capacity-achieving ensembles of punctured codes on graphs are the irregular repeat-accumulate (IRA) codes and accumulate-repeat-accumulate (ARA) codes. Recently, it was shown by Pfister et al. that properly designed nonsystematic IRA codes achieve

$\pi^{(0)}(x)$ (puncturing pattern)	Design rate	Capacity limit	Lower bound (ML decoding)	Iterative (IT) Decoding	Fractional gap to capacity (ML vs. IT)
0	0.500	0.187 dB	0.234 dB	0.299 dB	$\geq 41.5\%$
$0.102040x+0.06497x^2+0.06549x^5+0.00331x^6+0.39377x^{19}$	0.555	0.450 dB	0.473 dB	0.599 dB	$\geq 15.4\%$
$0.226410x+0.14149x^2+0.21268x^5+0.00001x^6+0.4424x^{19}$	0.625	0.816 dB	0.841 dB	1.028 dB	$\geq 11.9\%$
$0.348940x+0.21015x^2+0.38902x^5+0.00003x^6+0.48847x^{19}$	0.714	1.368 dB	1.398 dB	1.699 dB	$\geq 8.9\%$
$0.410320x+0.24330x^2+0.48388x^5+0.00004x^6+0.50541x^{19}$	0.769	1.777 dB	1.811 dB	2.215 dB	$\geq 7.8\%$
$0.469100x+0.28408x^2+0.56178x^5+0.00002x^6+0.53412x^{19}$	0.833	2.362 dB	2.404 dB	3.004 dB	$\geq 6.6\%$
$0.533750x+0.30992x^2+0.66375x^5+0.00001x^6+0.54837x^{19}$	0.909	3.343 dB	3.410 dB	4.634 dB	$\geq 5.2\%$

Table 3.2: Comparison of thresholds for ensembles of IP-LDPC codes where the original LDPC ensemble before puncturing has the degree distributions $\lambda(x) = 0.23403x + 0.21242x^2 + 0.14690x^5 + 0.10284x^6 + 0.30381x^{19}$ and $\rho(x) = 0.71875x^7 + 0.28125x^8$ (so its design rate is equal to $\frac{1}{2}$). The transmission of these codes takes place over a binary-input AWGN channel. The table compares values of $\frac{E_b}{N_0}$ referring to the capacity limit, the bound given in Theorem 3.3 (which provides a lower bound on $\frac{E_b}{N_0}$ under ML decoding), and thresholds under iterative message-passing decoding. The fractional gap to capacity (see the rightmost column) measures the ratio of the gap to capacity under optimal ML decoding and the achievable gap to capacity under (sub-optimal) iterative decoding. The pair of degree distributions for the ensemble of LDPC codes, and the polynomials which correspond to the puncturing patterns are taken from [35, Table 3].

$\pi^{(0)}(x)$ (puncturing pattern)	Design rate	Capacity limit	Lower bound (ML decoding)	Iterative (IT) Decoding	Fractional gap to capacity (ML vs. IT)
0	0.100	-1.286 dB	-1.248 dB	-1.028 dB	$\geq 14.5\%$
$0.486490x+0.69715x^2+0.03287x^3+0.04248x^4+0.69048x^7+0.45209x^{24}$	0.203	-0.953 dB	-0.917 dB	-0.731 dB	$\geq 16.3\%$
$0.655580x+0.83201x^2+0.48916x^3+0.33917x^4+0.63990x^7+0.76947x^{24}$	0.304	-0.605 dB	-0.570 dB	-0.317 dB	$\geq 12.0\%$
$0.745690x+0.87184x^2+0.38179x^3+0.48427x^4+0.74655x^7+0.79130x^{24}$	0.406	-0.226 dB	-0.189 dB	$+0.029$ dB	$\geq 14.7\%$
$0.838470x+0.65105x^2+0.04527x^3+0.95233x^4+0.74808x^7+0.80845x^{24}$	0.487	$+0.130$ dB	$+0.171$ dB	$+0.599$ dB	$\geq 8.7\%$
$0.979320x+0.46819x^2+0.71050x^3+0.59816x^4+0.79485x^7+0.05765x^{24}$	0.577	$+0.556$ dB	$+0.840$ dB	$+1.152$ dB	$\geq 47.7\%$
$0.895200x+0.84401x^2+0.98541x^3+0.42518x^4+0.92976x^7+0.30225x^{24}$	0.663	$+1.039$ dB	$+1.232$ dB	$+1.806$ dB	$\geq 25.2\%$
$0.910960x+0.91573x^2+0.23288x^3+0.40977x^4+0.99811x^7+0.15915x^{24}$	0.747	$+1.605$ dB	$+1.958$ dB	$+2.637$ dB	$\geq 34.2\%$
$0.904130x+0.96192x^2+0.35996x^3+0.96980x^4+0.31757x^7+0.89250x^{24}$	0.828	$+2.303$ dB	$+2.505$ dB	$+3.863$ dB	$\geq 13.0\%$

Table 3.3: Comparison of thresholds for ensembles of IP-LDPC codes where the original ensemble before puncturing has the degree distributions $\lambda(x) = 0.414936x + 0.183492x^2 + 0.013002x^3 + 0.093081x^4 + 0.147017x^7 + 0.148472x^{24}$ and $\rho(x) = 0.4x^2 + 0.6x^3$ (so its design rate is equal to $\frac{1}{10}$). The transmission of these codes takes place over a binary-input AWGN channel. The table compares values of $\frac{E_b}{N_0}$ referring to the capacity limit, the bound given in Theorem 3.3 (which provides a lower bound on $\frac{E_b}{N_0}$ under ML decoding), and thresholds under iterative message-passing decoding. The fractional gap to capacity (see the rightmost column) measures the ratio of the gap to capacity under optimal ML decoding and the achievable gap to capacity under (sub-optimal) iterative decoding. The pair of degree distributions for the ensemble of LDPC codes, and the polynomials which correspond to the puncturing patterns are taken from [34, Table 5.1].

the capacity of the BEC with bounded decoding complexity per information bit [65]. This bounded complexity result is achieved by puncturing all the information bits of the IRA codes, and allowing in this way a sufficient number of state nodes in the Tanner graph representing the codes. This is in contrast to all previous constructions of capacity-achieving LDPC codes which refer to bipartite graphs without state nodes and whose complexity becomes unbounded as their gap to capacity vanishes (for an information-theoretic proof which explains why the complexity becomes unbounded in this case, the reader is referred to [81, Theorem 2.1]). The decoding complexity of punctured LDPC codes for parallel channels is addressed in the next section.

3.5 Lower Bounds on the Decoding Complexity of LDPC Codes for Parallel Channels

The scope of this section is to derive a lower bound on the decoding complexity of LDPC codes for parallel MBIOS channels. The lower bound holds under iterative message-passing decoding, and it grows like the logarithm of the inverse of the gap (in rate) to capacity. Interestingly, a logarithmic behavior of the parity-check density (which forms a measure of the decoding complexity per iteration) in terms of the gap to capacity also characterizes the upper bound derived in [50, Section 3]; this upper bound refers to MacKay's ensemble of LDPC codes whose transmission takes place over a set of parallel MBIOS channels.

In the previous section we regarded the transmission of punctured LDPC codes over MBIOS channels as a special case of the transmission of the original codes (before puncturing) over a set of parallel MBIOS channels. Hence, the aforementioned bound is later applied to obtain lower bounds on the decoding complexity of punctured LDPC codes. This section refers to an appendix which suggests a simplified re-derivation of [65, Theorems 3 and 4], and shows that the bounds introduced in this section are tighter.

3.5.1 A Lower Bound on the Decoding Complexity for Parallel MBIOS Channels

Consider a binary linear block code which is represented by a bipartite graph, and assume that the graph serves for the decoding with an iterative algorithm. Following [50] and [65], the decoding complexity under message-passing decoding is defined as the number of edges in the graph normalized per information bit. This quantity measures the number of messages which are delivered through the edges of the graph

(from left to right and vice versa) during a single iteration. Equivalently, since there is a one-to-one correspondence between a bipartite graph and the parity-check matrix H which represents the code, the decoding complexity is also equal to the number of non-zero elements in H normalized per information bit (i.e., it is equal to the density of the parity-check matrix [81, Definition 2.2]). Hence, the decoding complexity and performance of iteratively decoded binary linear block codes depend on the specific representation of the code by a parity-check matrix. Since the average right degree (a_{R}) of a bipartite graph is equal to the number of edges per parity-check equation, then the average right degree and the decoding complexity are related quantities. Consider an ensemble of LDPC codes whose design rate is R_{d}. It is natural to relate the decoding complexity of the ensemble, say χ_D, to its average right degree and design rate, as follows:

$$\chi_{\mathrm{D}} = \left(\frac{1 - R_{\mathrm{d}}}{R_{\mathrm{d}}}\right) a_{\mathrm{R}} \,.$$

We note that a_{R} is fixed for all the codes from an ensemble of LDPC codes with a given pair of degree distributions.

The following lemma is used in the continuation for the derivation of a lower bound on the decoding complexity per iteration under message-passing decoding.

Lemma 3.8 Let Γ be the right degree distribution of an ensemble of LDPC codes. Then

$$\Gamma(\alpha) \geq \alpha^{a_{\mathrm{R}}}, \quad \forall\, \alpha \geq 0.$$

Proof: Using the convexity of the function $f(x) = \alpha^x$, it follows from Jensen's inequality that for $\alpha \geq 0$

$$\Gamma(\alpha) = \sum_{i=1}^{\infty} \Gamma_i \alpha^i \geq \alpha^{\sum_{i=1}^{\infty} i\Gamma_i} = \alpha^{a_{\mathrm{R}}}.$$

■

Consider a sequence of ensembles of LDPC codes, $\{(n_r, \lambda_{\mathbf{r}}, \rho)\}_{r=1}^{\infty}$, whose transmission takes place over a set of J statistically independent parallel MBIOS channels. Let C_j and p_j be the capacity and the fraction of code bits assigned to the j^{th} channel, respectively (where $j \in \{1, \ldots, J\}$). We define the average capacity of the set of J parallel channels as $\overline{C} \triangleq \sum_{j=1}^{J} p_j C_j$. For an ensemble of LDPC codes which achieves vanishing bit error probability as the block length goes to infinity, the multiplicative gap (in rate) to capacity is defined as

$$\varepsilon \triangleq 1 - \frac{R_{\mathrm{d}}}{\overline{C}}. \tag{3.58}$$

We now present a lower bound on the decoding complexity per iteration under message-passing decoding for this sequence. The bound is given in terms of the gap to capacity.

Theorem 3.4 Let a sequence of ensembles of LDPC codes, $\{(n_r, \lambda_{\mathbf{r}}, \rho)\}_{r=1}^{\infty}$, be transmitted over a set of J statistically independent parallel MBIOS channels. Let C_j be the capacity of the j^{th} channel (where $j \in \{1, \ldots, J\}$), and denote the average capacity by $\overline{C} \triangleq \sum_{j=1}^{J} p_j C_j$. If this sequence achieves a fraction $1 - \varepsilon$ of $\overline{C}$ with vanishing bit error probability, then the asymptotic decoding complexity under message-passing decoding satisfies

$$\chi_{\mathrm{D}}(\varepsilon) \geq K_1 + K_2 \ln\left(\frac{1}{\varepsilon}\right). \tag{3.59}$$

The coefficients $K_{1,2}$ in this lower bound are as follows:

$$K_1 = -\frac{(1-\overline{C}) \ln\left(\frac{1}{2\ln 2}\frac{1-\overline{C}}{\overline{C}}\right)}{\overline{C} \ln\left(\sum_{j=1}^{J} q_j g_{j,1}\right)}, \quad K_2 = -\frac{1-\overline{C}}{\overline{C} \ln\left(\sum_{j=1}^{J} q_j g_{j,1}\right)} \tag{3.60}$$

where $g_{j,1}$ is introduced in (3.3), and q_j is introduced in (3.26) and is assumed to be positive for all $j \in \{1, \ldots, J\}$. For parallel BECs, the term $\frac{1}{2\ln 2}$ can be removed from the numerator in the expression of K_1.

Proof: Substituting (3.58) in (3.27) gives

$$(1-\varepsilon)\overline{C} \leq 1 - \frac{1-\overline{C}}{1 - \frac{1}{2\ln 2}\sum_{p=1}^{\infty}\left\{\frac{1}{p(2p-1)}\Gamma\left(\sum_{j=1}^{J} q_j g_{j,p}\right)\right\}}. \tag{3.61}$$

Since $g_{j,p}$ in (3.3) is non-negative for $j \in \{1, \ldots, J\}$ and $p \in \mathbb{N}$, and the function Γ is non-negative on $\mathbb{R}^+$, then the terms in the infinite sum above are all non-negative. By the truncation of this series where we only take its first term (note that this is the largest term in the sum), we obtain a lower bound on the RHS of (3.61). This implies that

$$(1-\varepsilon)\overline{C} \leq 1 - \frac{1-\overline{C}}{1 - \frac{1}{2\ln 2}\Gamma\left(\sum_{j=1}^{J} q_j g_{j,1}\right)}.$$

Invoking Lemma 3.8 yields that

$$(1-\varepsilon)\overline{C} \leq 1 - \frac{1-\overline{C}}{1 - \frac{1}{2\ln 2}\left(\sum_{j=1}^{J} q_j g_{j,1}\right)^{a_{\mathrm{R}}}}.$$

The solution of the last inequality for the average right degree (a_{R}) gives

$$\begin{aligned} a_{\mathrm{R}} &\geq -\frac{\ln\left(\frac{1}{2\ln 2}\left(1+\frac{1-\overline{C}}{\overline{C}\varepsilon}\right)\right)}{\ln\left(\sum_{j=1}^{J} q_j g_{j,1}\right)} \\ &> K_1' + K_2' \ln\left(\frac{1}{\varepsilon}\right) \end{aligned} \tag{3.62}$$

where the last step follows by dropping the 1 which appeared inside the logarithm at the numerator (this step is valid since the denominator is strictly negative), and

$$K_1' = -\frac{\ln\left(\frac{1}{2\ln 2}\frac{1-\overline{C}}{\overline{C}}\right)}{\ln\left(\sum_{j=1}^{J} q_j g_{j,1}\right)}, \quad K_2' = -\frac{1}{\ln\left(\sum_{j=1}^{J} q_j g_{j,1}\right)}.$$

Since $R_{\mathrm{d}} < \overline{C}$, it follows that $\chi_{\mathrm{D}} = \frac{1-R_{\mathrm{d}}}{R_{\mathrm{d}}}\, a_{\mathrm{R}} > \frac{1-\overline{C}}{\overline{C}}\, a_{\mathrm{R}}$. The proof of the lower bound on the decoding complexity for parallel MBIOS channels follows by multiplying both sides of (3.62) by $\frac{1-\overline{C}}{\overline{C}}$.

For parallel BECs, we get from (3.3) that for every $p \in \mathbb{N}$

$$g_{j,p} = \int_0^\infty a(l;j)(1+e^{-l}) \tanh^{2p}\left(\frac{l}{2}\right) \mathrm{d}l = 1 - \varepsilon_j$$

where ε_j denotes the erasure probability of the j^{th} BEC. This gives

$$\begin{aligned} &\frac{1}{2\ln 2} \sum_{p=1}^{\infty} \left\{ \frac{1}{p(2p-1)} \Gamma\left(\sum_{j=1}^{J} q_j g_{j,p}\right) \right\} \\ &= \frac{1}{2\ln 2} \sum_{p=1}^{\infty} \frac{1}{p(2p-1)} \cdot \Gamma\left(\sum_{j=1}^{J} q_j g_{j,1}\right) \\ &= \Gamma\left(\sum_{j=1}^{J} q_j g_{j,1}\right). \end{aligned}$$

Substituting this in (3.61), gives

$$(1-\varepsilon)\overline{C} \leq 1 - \frac{1-\overline{C}}{1-\Gamma\left(\sum_{j=1}^{J} q_j g_{j,1}\right)}.$$

The continuation of the proof follows the same steps as the proof for parallel MBIOS channels, and leads to an improved coefficient K_1 where the factor $\frac{1}{2\ln 2}$ in the numerator of K_1 for general MBIOS channels (see (3.60)) is replaced by 1. ■

We proceed the analysis by the derivation of lower bounds on the decoding complexity of sequences of ensembles of punctured LDPC codes where it is assumed that these sequences achieve vanishing bit error probability; similarly to Theorem 3.4, the lower bounds are expressed in terms of the multiplicative gap (in rate) to capacity.

3.5.2 Lower Bounds on the Decoding Complexity for Punctured LDPC Codes

As discussed in the previous section, transmission of punctured codes can be interpreted as a special case of transmitting the original (un-punctured) codes over a set of parallel channels where these component channels are formed by a mixture of the communication channel and BECs whose erasure probabilities are the puncturing rates of the different subsets of code bits. Hence, the bounds on the decoding complexity of punctured codes can be derived as special cases of the bound given in Theorem 3.4. For the sake of brevity, we derive these bounds by using the upper bounds on the achievable rates of punctured LDPC codes as given in Theorem 3.2 (for random puncturing) and Theorem 3.3 (for intentional puncturing). Note that the derivation of these two theorems relies on Theorem 3.1 (as shown in Figure 3.1 on p. 121).

Consider an ensemble of LDPC codes of length n and design rate R'_{d}, and let the code bits be partitioned into J disjoint sets where the j^{th} set contains a fraction p_j of these bits ($j \in \{1, \ldots, J\}$). Assume that the bits in the j^{th} set are randomly punctured at rate π_j, and let the punctured codes be transmitted over an MBIOS channel whose capacity is C. As shown in the previous section, this is equivalent to transmitting the original (un-punctured) codes over a set of J parallel channels, where the j^{th} set of code bits is transmitted over a channel whose capacity is $C_j = (1-\pi_j)C$. The average capacity of this set of J parallel channels is therefore given by

$$\overline{C} = \sum_{j=1}^{J} p_j \, (1-\pi_j) C = \big(1 - \sum_{j=1}^{J} p_j \pi_j\big) C = (1-\gamma) C \tag{3.63}$$

where $\gamma \triangleq \sum_{j=1}^{J} p_j \pi_j$ is the overall puncturing rate. Denote the design rate of the punctured codes by $R_{\mathrm{d}} \triangleq \frac{R'_{\mathrm{d}}}{1-\gamma}$ (see Definition 3.1 on p. 102), then it follows that the multiplicative gap to capacity of the punctured codes is given by

$$\varepsilon = 1 - \frac{R_{\mathrm{d}}}{C} = 1 - \frac{R'_{\mathrm{d}}}{\overline{C}}. \tag{3.64}$$

For punctured codes, the iterative decoder is based on the bipartite graph of the 'mother code' where the channel input to the variable nodes which correspond

to the punctured code bits is defined to be 0. Hence, the decoding complexity of the punctured ensemble under message-passing decoding is identical to the decoding complexity of the original ensemble (before puncturing), and is given by

$$\begin{aligned} \chi_{\mathrm{D}} &= \left(\frac{1-R'_{\mathrm{d}}}{R'_{\mathrm{d}}}\right) a_{\mathrm{R}} \\ &= \left(\frac{1-(1-\gamma)R_{\mathrm{d}}}{(1-\gamma)R_{\mathrm{d}}}\right) a_{\mathrm{R}}\,. \end{aligned} \tag{3.65}$$

In the following, we derive a lower bound on the decoding complexity of a sequence of ensembles of RP-LDPC codes.

Theorem 3.5 Let $\{(n_r, \lambda, \rho)\}_{r=1}^{\infty}$ be a sequence of ensembles of LDPC codes whose block length (n_r) tends to infinity as $r \to \infty$. Assume that a sequence of ensembles of RP-LDPC codes is constructed in the following way: for each code from an ensemble of the original sequence, a subset of αn_r code bits is a-priori selected, and these bits are randomly punctured at a fixed rate (P_{pct}). Assume that the punctured codes are transmitted over an MBIOS channel with capacity C, and that as r tends to infinity, the sequence of ensembles of punctured codes achieves a fraction $1-\varepsilon$ of the capacity with vanishing bit error probability. Then in probability 1 w.r.t. the random puncturing patterns, the decoding complexity of this sequence under message-passing decoding satisfies

$$\chi_{\mathrm{D}}(\varepsilon) \geq K_1 + K_2 \ln\left(\frac{1}{\varepsilon}\right). \tag{3.66}$$

The coefficients $K_{1,2}$ in this lower bound are as follows:

$$K_1 = -\frac{(1-\overline{C})\,\ln\left(\frac{1}{2\ln 2}\frac{1-\overline{C}}{\overline{C}}\right)}{\overline{C}\,\ln\big((1-P_{\mathrm{pct}}+\xi)g_1\big)}, \quad K_2 = -\frac{1-\overline{C}}{\overline{C}\,\ln\big((1-P_{\mathrm{pct}}+\xi)g_1\big)} \tag{3.67}$$

where g_1 is introduced in (3.41), ξ is introduced in (3.43), and $\overline{C} \triangleq (1-\alpha P_{\mathrm{pct}})C$. For the particular case of a BEC, the term $\frac{1}{2\ln 2}$ can be dropped, thus improving the tightness of the additive term (K_1) in the lower bound.

Proof: Since the code bits of a subset of the code bits whose size is αn_r are randomly punctured at rate P_{pct}, then the average puncturing rate is given by $\gamma = \alpha P_{\mathrm{pct}}$. Hence, Eq. (3.63) yields that $\overline{C} = (1-\alpha P_{\mathrm{pct}})C$. By multiplying both sides of (3.42) by $1-\alpha P_{\mathrm{pct}}$ and getting from (3.64) that $R_{\mathrm{d}} = (1-\varepsilon)C$, we obtain

$$(1-\varepsilon)\overline{C} \leq 1 - \frac{1-\overline{C}}{1-\dfrac{1}{2\ln 2}\displaystyle\sum_{p=1}^{\infty}\left\{\frac{1}{p(2p-1)}\,\Gamma\Big((1-P_{\mathrm{pct}}+\xi)g_p\Big)\right\}}.$$

Following the same steps as in the proof of Theorem 3.4, we get a lower bound on the average right degree of the bipartite graph which corresponds to the pair of degree distributions (λ, ρ). This lower bound is of the form

$$a_{\mathrm{R}} > K_1' + K_2' \ln\left(\frac{1}{\varepsilon}\right) \tag{3.68}$$

where

$$K_1' = -\frac{\ln\left(\frac{1}{2\ln 2}\frac{1-\overline{C}}{\overline{C}}\right)}{\ln\Big((1-P_{\mathrm{pct}}+\xi)g_1\Big)}, \quad K_2' = -\frac{1}{\ln\Big((1-P_{\mathrm{pct}}+\xi)g_1\Big)}.$$

Note that K_2 is positive; this follows from (3.43), which yields that $\xi < (1-\alpha)P_{\mathrm{pct}}$ (due to the fact that the integral of λ over the interval [0, 1] is upper bounded by $\frac{1}{2}$). This assures that as the gap (in rate) to capacity vanishes, the lower bound on a_{R} scales like the logarithm of the inverse of this gap.

From (3.64), we get $R_{\mathrm{d}}' = (1-\varepsilon)\overline{C} < \overline{C}$, and therefore $\chi_{\mathrm{D}} = \frac{1-R_{\mathrm{d}}'}{R_{\mathrm{d}}'}\, a_{\mathrm{R}} > \frac{1-\overline{C}}{\overline{C}}\, a_{\mathrm{R}}$. The proof of the lower bound on the decoding complexity is completed by multiplying both sides of (3.68) by $\frac{1-\overline{C}}{\overline{C}}$. In the particular case where the communication channel is a BEC, following the same concept as in the proof of Theorem 3.4 leads to the improved coefficient K_1. ■

The upper bound on the decoding complexity for sequences of ensembles of IP-LDPC codes is also given in terms of the gap between the rate of the punctured rate and the channel capacity.

Theorem 3.6 Let $\{(n_r, \lambda, \rho, \pi^{(0)})\}_{r=1}^{\infty}$ be a sequence of ensembles of IP-LDPC codes transmitted over an MBIOS channel whose capacity is C. If this sequence achieves a fraction $1-\varepsilon$ of the capacity with vanishing bit error probability, then in probability 1 w.r.t. the random puncturing patterns, the decoding complexity of this sequence under iterative message-passing decoding satisfies

$$\chi_{\mathrm{D}}(\varepsilon) \geq K_1 + K_2 \ln\left(\frac{1}{\varepsilon}\right). \tag{3.69}$$

The coefficients $K_{1,2}$ in this lower bound are as follows:

$$K_1 = -\frac{(1-\overline{C})\,\ln\left(\frac{1}{2\ln 2}\frac{1-\overline{C}}{\overline{C}}\right)}{\overline{C}\,\ln\left(\left(1-\sum_{j=1}^{\infty}\lambda_j\pi_j\right)g_p\right)}, \quad K_2 = -\frac{1-\overline{C}}{\overline{C}\,\ln\left(\left(1-\sum_{j=1}^{\infty}\lambda_j\pi_j\right)g_p\right)} \tag{3.70}$$

where g_1 is introduced in (3.41), and $\overline{C} \triangleq (1-\sum_{j=1}^{\infty}\Lambda_j\pi_j)C$. For the particular case of a BEC, the term $\frac{1}{2\ln 2}$ can be dropped, thus improving the tightness of the additive term (K_1) in the lower bound.

Proof: The proof follows from the same concepts as the proof of Theorem 3.5, but is based on (3.51) instead of (3.42). Note that K_2, which reflects the logarithmic growth rate of the lower bound in (3.69), is always positive; this follows from (3.70) and due to the fact that from (3.41), $g_1 < 1$, and also $0 < 1 - \sum_{j=1}^{\infty} \lambda_j \pi_j \leq 1$. ■

3.5.3 Re-Derivation of Reported Lower Bounds on the Decoding Complexity

In [65, Theorems 3 and 4], Pfister et al. introduced lower bounds on the decoding complexity of punctured codes under iterative decoding. The bounds were derived for the case where a subset of linearly independent code bits whose size is equal to the code dimension are randomly punctured at a fixed rate (P_{pct}), and the transmission of the codes takes place over an MBIOS channel. In particular, this scenario corresponds to RP-LDPC codes (see Section 3.4.2) where we choose a subset of the code bits to be randomly punctured at rate P_{pct}; under the assumption in [65, Theorems 3 and 4], the fraction (α) of the code bits which are randomly punctured is equal to the code rate. In the appendix, we show that for RP-LDPC codes, the lower bounds on the decoding complexity given in [65, Theorems 3 and 4] follow from a looser version of the bound in Theorem 3.5.

3.6 Summary and Outlook

The main result in this paper, Theorem 3.1, provides an upper bound on the asymptotic rate of a sequence of ensembles of low-density parity-check (LDPC) codes which achieves vanishing bit error probability. We assume that the communication takes place over a set of parallel memoryless binary-input output-symmetric (MBIOS) channels. The derivation of Theorem 3.1 relies on upper and lower bounds on the conditional entropy of the transmitted codeword given the received sequence at the output of the parallel channels (see Section 3.2), and it is valid under optimal maximum-likelihood (ML) decoding (or any sub-optimal decoding algorithm). This theorem enables the derivation of a lower bound on the decoding complexity (per iteration) of ensembles of LDPC codes under message-passing iterative decoding when the transmission of the codes takes place over parallel MBIOS channels. The latter bound is given in terms of the gap between the rate of these codes for which reliable communication is achievable and the channel capacity. Similarly to a lower bound on the decoding complexity of ensembles of LDPC codes for a single MBIOS channel [81], the lower bound on the decoding complexity which is derived for parallel channels

also grows like the log of the inverse of the gap to capacity.

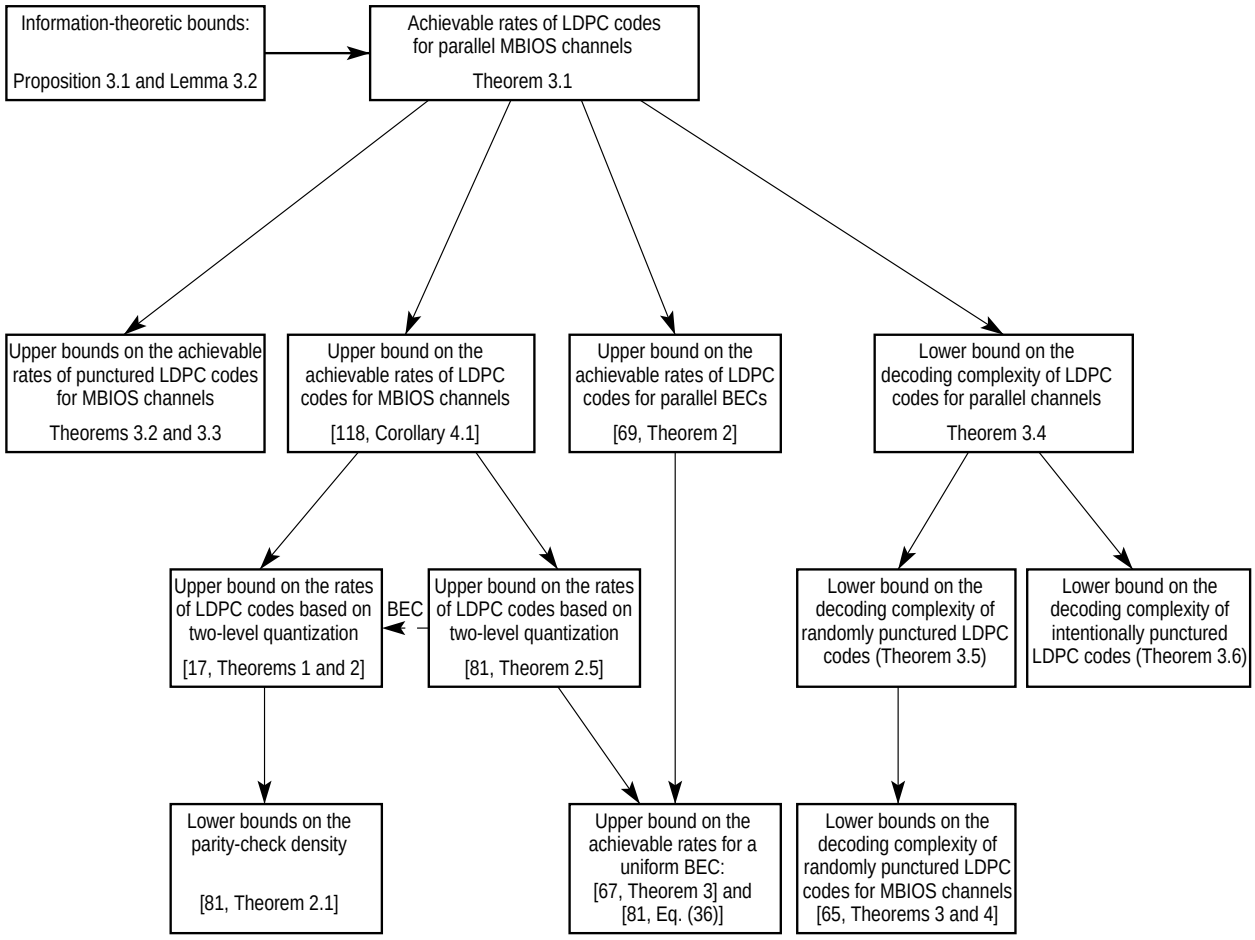

Figure 3.1: An interconnections diagram among the bounds in this paper and some previously reported bounds which follow as special cases.

Theorem 3.1 can be used for various applications which form particular cases of communication over parallel channels, e.g., intentionally punctured LDPC codes [35], non-uniformly error protected LDPC codes [69], and LDPC-coded modulation (see e.g., [37, 112]). In Section 3.4, we use Theorem 3.1 for the derivation of upper bounds on the achievable rates under ML decoding of (randomly and intentionally) punctured LDPC codes whose transmission takes place over an MBIOS channel. It is exemplified numerically that for various good ensembles of intentionally punctured LDPC codes, the asymptotic loss in performance due to the code structure is still non-negligible as compared to the corresponding loss due to the sub-optimality of iterative decoding (as compared to optimal ML decoding). Looser versions of the bounds derived in

this paper for punctured LDPC codes suggest a simplified re-derivation of previously reported bounds on the decoding complexity of randomly punctured LDPC codes (see [65, Theorems 3 and 4]).

Interconnections between the theorems introduced in this paper and some previously reported results which follow as special cases are shown in Figure 3.1.

Appendix

3.1 Re-derivation of [65, Theorems 3 and 4]

In the following, we start with the re-derivation of [65, Theorem 4] for general MBIOS channels, and then re-derive the refined bound in [65, Theorem 3] for a BEC. For the re-derivation of [65, Theorems 3 and 4] we rely on Theorem 3.5 whose derivation is based on Theorem 3.2. Hence, we first loosen the upper bound on the achievable rates given in (3.42), and then re-derive [65, Theorem 4] as a consequence of this looser version. The loosening of (3.42) is done by replacing the positive parameter ξ introduced in (3.43) by zero, and then using the lower bound on Γ from Lemma 3.8. This gives

$$R_{\mathrm{d}} \leq \frac{1}{1-\alpha P_{\mathrm{pct}}} \left(1 - \frac{1-(1-\alpha P_{\mathrm{pct}})C}{1-\frac{1}{2\ln 2}\sum_{p=1}^{\infty}\left\{\frac{1}{p(2p-1)}\Big((1-P_{\mathrm{pct}})g_p\Big)^{a_{\mathrm{R}}}\right\}} \right). \tag{3.A.1}$$

Finally, truncating the infinite series in the RHS of (3.A.1) by only taking its first term which corresponds to $p=1$ further loosens the upper bound on the achievable rates, and gives

$$R_{\mathrm{d}} \leq \frac{1}{1-\alpha P_{\mathrm{pct}}} \left(1 - \frac{1-(1-\alpha P_{\mathrm{pct}})C}{1-\frac{1}{2\ln 2}\Big((1-P_{\mathrm{pct}})g_1\Big)^{a_{\mathrm{R}}}} \right). \tag{3.A.2}$$

From (3.64), we get the inequality

$$(1-\varepsilon)(1-\alpha P_{\mathrm{pct}})C \leq 1 - \frac{1-(1-\alpha P_{\mathrm{pct}})C}{1-\frac{1}{2\ln 2}\Big((1-P_{\mathrm{pct}})g_1\Big)^{a_{\mathrm{R}}}}.$$

which after straightforward algebra gives

$$1 + \frac{1-(1-\alpha P_{\mathrm{pct}})C}{\varepsilon C(1-\alpha P_{\mathrm{pct}})} \leq 2\ln 2 \left(\frac{1}{(1-P_{\mathrm{pct}})g_1}\right)^{a_{\mathrm{R}}}. \tag{3.A.3}$$

We proceed by giving a simple lower bound on g_1.

Lemma 3.A.1 For g_1 introduced in (3.41), the following inequality holds

$$g_1 \geq (1-2w)^2$$

where

$$w \triangleq P_{\rm e}(a) = \frac{1}{2}\Pr(L=0) + \int_{-\infty}^{0^-} a(l)\,{\rm d}l$$

designates the uncoded bit error probability of the MBIOS channel given the channel input is 1.

Proof: Based on the symmetry property where $a(l) = e^l a(-l)$ and Jensen's inequality, we get

$$\begin{aligned}
g_1 &= \int_0^\infty a(l)\,(1+e^{-l})\tanh^2\left(\frac{l}{2}\right){\rm d}l \\
&= \int_{-\infty}^\infty a(l)\tanh^2\left(\frac{l}{2}\right){\rm d}l \\
&\geq \left(\int_{-\infty}^\infty a(l)\tanh\left(\frac{l}{2}\right){\rm d}l\right)^2 \\
&= \left(\int_0^\infty a(l)\,(1+e^{-l})\tanh\left(\frac{l}{2}\right){\rm d}l\right)^2 \\
&= \left(\int_0^\infty a(l)\,(1-e^{-l}){\rm d}l\right)^2 \\
&= \left(\int_{0^+}^\infty \big(a(l)-a(-l)\big){\rm d}l\right)^2 \\
&= \left(1-\Pr(L=0)-2\int_{-\infty}^{0^-} a(l)\,{\rm d}l\right)^2 \\
&= (1-2w)^2.
\end{aligned}$$

■

Replacing g_1 in the RHS of (3.A.3) by its lower bound from Lemma 3.A.1 gives

$$\begin{aligned}
1+\frac{1-(1-\alpha P_{\rm pct})C}{\varepsilon C(1-\alpha P_{\rm pct})} &\leq 2\ln 2\left(\frac{1}{(1-P_{\rm pct})(1-2w)^2}\right)^{a_{\rm R}} \\
&\leq 2\ln 2\left(\frac{1}{(1-P_{\rm pct})(1-2w)}\right)^{2a_{\rm R}}.
\end{aligned}$$

Solving the last inequality for $a_{\rm R}$ gives

$$a_{\rm R} \geq \frac{\ln\left(\frac{1}{2\ln 2}\left(1+\frac{1-(1-\alpha P_{\rm pct})C}{\varepsilon C(1-\alpha P_{\rm pct})}\right)\right)}{2\ln\left(\frac{1}{(1-P_{\rm pct})(1-2w)}\right)}.$$

Based on the equality (3.65) which relates the complexity under message-passing decoding to the average right degree (a_{R}) and since $R_{\mathrm{d}} < C$, we get from the last inequality

$$\chi_{\mathrm{D}}(\varepsilon) \geq \frac{1-C}{2C} \frac{\ln\left(\frac{1}{2\ln 2}\left(1+\frac{1-(1-\alpha P_{\mathrm{pct}})C}{\varepsilon C(1-\alpha P_{\mathrm{pct}})}\right)\right)}{\ln\left(\frac{1}{(1-P_{\mathrm{pct}})(1-2w)}\right)}. \tag{3.A.4}$$

Note that $\alpha = R'_{\mathrm{d}}$ in [65, Theorem 4]. This gives the equality $\alpha = (1-\varepsilon)\overline{C} = (1-\varepsilon)(1-\alpha P_{\mathrm{pct}})C$ whose solution is

$$\alpha = \frac{(1-\varepsilon)C}{1+(1-\varepsilon)CP_{\mathrm{pct}}}. \tag{3.A.5}$$

Finally, the substitution of α in (3.A.5) into the RHS of (3.A.4) gives

$$\begin{aligned}
\chi_{\mathrm{D}}(\varepsilon) &\geq \frac{1-C}{2C} \frac{\ln\left(\frac{1}{2\ln 2}\left(1+\frac{1-(1-P_{\mathrm{pct}})C-\varepsilon CP_{\mathrm{pct}}}{\varepsilon C}\right)\right)}{\ln\left(\frac{1}{(1-P_{\mathrm{pct}})(1-2w)}\right)} \\
&\geq \frac{1-C}{2C} \frac{\ln\left(\frac{1}{\varepsilon}\frac{1-(1-P_{\mathrm{pct}})C}{2C\ln 2}\right)}{\ln\left(\frac{1}{(1-P_{\mathrm{pct}})(1-2w)}\right)}.
\end{aligned} \tag{3.A.6}$$

which coincides with [65, Theorem 4] for a sequence of ensembles of randomly punctured LDPC codes.

For the derivation of the refined bound for the BEC which is given in [65, Theorem 3], we start from (3.A.1). The refinement of the latter bound is due to the fact that for the BEC, g_p in (3.41) is independent of p, and is equal to $g_p = 1 - P_{\mathrm{BEC}}$ where P_{BEC} designates the erasure probability of the BEC. From (3.A.1), we get the following upper bound on the achievable rates:

$$R_{\mathrm{d}} \leq \frac{1}{1-\alpha P_{\mathrm{pct}}} \left(1 - \frac{1-(1-\alpha P_{\mathrm{pct}})C}{1-\Big((1-P_{\mathrm{pct}})(1-P_{\mathrm{BEC}})\Big)^{a_{\mathrm{R}}}}\right)$$

which follows from the equality $\sum_{p=1}^{\infty} \frac{1}{2p(2p-1)} = \ln 2$. Substituting $R_{\mathrm{d}} = (1-\varepsilon)(1-P_{\mathrm{BEC}})$ and the α in (3.A.5) gives a lower bound on a_{R}. Finally, the lower bound in [65, Theorem 3] follows from the resulting lower bound on a_{R} and the inequality $\chi_D(\varepsilon) \geq \frac{1-C}{C} a_{\mathrm{R}}$.

Chapter 4

Bounds on the Number of Iterations for Turbo-Like Ensembles over the Binary Erasure Channel

This chapter is a preprint of

- I. Sason and G. Wiechman, "Bounds on the number of iterations for turbo-like ensembles over the binary erasure channel," submitted to *IEEE Trans. on Information Theory*, November 2007.

Chapter Overview: This paper provides simple lower bounds on the number of iterations which is required for successful message-passing decoding of some important families of graph-based code ensembles (including low-density parity-check codes and variations of repeat-accumulate codes). The transmission of the code ensembles is assumed to take place over a binary erasure channel, and the bounds refer to the asymptotic case where we let the block length tend to infinity. The simplicity of the bounds derived in this paper stems from the fact that they are easily evaluated and are expressed in terms of some basic parameters of the ensemble which include the fraction of degree-2 variable nodes, the target bit erasure probability and the gap between the channel capacity and the design rate of the ensemble. This paper demonstrates that the number of iterations which is required for successful message-passing decoding scales at least like the inverse of the gap (in rate) to capacity, provided that the fraction of degree-2 variable nodes of these turbo-like ensembles does not vanish (hence, the number of iterations becomes unbounded as the gap to capacity vanishes).

4.1 Introduction

During the last decade, there have been many developments in the construction and analysis of low-complexity error-correcting codes which closely approach the Shannon capacity limit of many standard communication channels with feasible complexity. These codes are understood to be codes defined on graphs, together with the associated iterative decoding algorithms. Graphs serve not only to describe the codes themselves, but more importantly, they structure the operation of their efficient sub-optimal iterative decoding algorithms.

Proper design of codes defined on graphs enables to asymptotically achieve the capacity of the binary erasure channel (BEC) under iterative message-passing decoding. Capacity-achieving sequences of ensembles of low-density parity-check (LDPC) codes were originally introduced by Shokrollahi [94] and by Luby et al. [51], and a systematic study of capacity-achieving sequences of LDPC ensembles was presented by Oswald and Shokrollahi [63] for the BEC. Analytical bounds on the maximal achievable rates of LDPC ensembles were derived by Barak et al. [10] for the asymptotic case where the block length tends to infinity; this analysis provides a lower bound on the gap between the channel capacity and the achievable rates of LDPC ensembles under iterative decoding. The decoding complexity of LDPC codes under iterative message-passing decoding scales linearly with the block length, though their encoding complexity is in general super-linear with the block length; this motivated the introduction of repeat-accumulate codes and their more recent variants (see, e.g., [4], [40] and [64]) whose encoding and decoding complexities under iterative message-passing decoding are both inherently linear with the block length. Due to the simplicity of the density evolution analysis for the BEC, suitable constructions of capacity-achieving ensembles of variants of repeat-accumulate codes were devised in [40], [65], [64] and [80]. All these works rely on the density evolution analysis of codes defined on graphs for the BEC, and provide an asymptotic analysis which refers to the case where we let the block length of these code ensembles tend to infinity. Another innovative coding technique, introduced by Shokrollahi [95], enables to achieve the capacity of the BEC with encoding and decoding complexities which scale linearly with the block length, and it has the additional pleasing property of achieving the capacity without the knowledge of the erasure probability of the channel.

The performance analysis of finite-length LDPC code ensembles whose transmission takes place over the BEC was introduced by Di et al. [22]. This analysis considers sub-optimal iterative message-passing decoding as well as optimal maximum-likelihood decoding. In [6], an efficient approach to the design of LDPC codes of

finite length was introduced by Amraoui et al.; this approach is specialized for the BEC, and it enables to design such code ensembles which perform well under iterative decoding with a practical constraint on the block length. In [72], Richardson and Urbanke initiated the analysis of the distribution of the number of iterations needed for the decoding of LDPC ensembles of finite block length which are communicated over the BEC.

For general channels, the number of iterations is an important factor in assessing the decoding complexity of graph-based codes under iterative message-passing decoding. The second factor determining the decoding complexity of such codes is the complexity of the Tanner graph which is used to represent the code; this latter quantity, defined as the number of edges in the graph per information bit, serves as a measure for the decoding complexity per iteration.

The extrinsic information transfer (EXIT) charts, pioneered by Stephan ten Brink [102, 101], form a powerful tool for an efficient design of codes defined on graphs by tracing the convergence behavior of their iterative decoders. EXIT charts provide a good approximative engineering tool for tracing the convergence behavior of soft-input soft-output iterative decoders; they suggest a simplified visualization of the convergence of these decoding algorithms, based on a single parameter which represents the exchange of extrinsic information between the constituent decoders. For the BEC, the EXIT charts coincide with the density evolution analysis (see [74]) which is simplified in this case to a one-dimensional analysis.

A numerical approach for the joint optimization of the design rate and decoding complexity of LDPC ensembles was provided in [8]; it is assumed there that the transmission of these code ensembles takes place over a memoryless binary-input output-symmetric (MBIOS) channel, and the analysis refers to the asymptotic case where we let the block length tend to infinity. For the simplification of the numerical optimization, a suitable approximation of the number of iterations was used in [8] to formulate this joint optimization as a convex optimization problem. Due to the efficient tools which currently exist for a numerical solution of convex optimization problems, this approach suggests an engineering tool for the design of good LDPC ensembles which possess an attractive tradeoff between the decoding complexity and the asymptotic gap to capacity (where the block length of these code ensembles is large enough). This numerical approach however is not amenable for drawing rigorous theoretical conclusions on the tradeoff between the number of iterations and the performance of the code ensembles. A different numerical approach for approximating the number of iterations for LDPC ensembles operating over the BEC is addressed in [52].

A different approach for characterizing the complexity of iterative decoders was suggested by Khandekar and McEliece (see [43, 42, 56]). Their questions and conjectures were related to the tradeoff between the asymptotic achievable rates and the complexity under iterative message-passing decoding; they initiated a study of the encoding and decoding complexity of graph-based codes in terms of the achievable gap (in rate) to capacity. It was conjectured there that for a large class of channels, if the design rate of a suitably designed ensemble forms a fraction $1-\varepsilon$ of the channel capacity, then the decoding complexity scales like $\frac{1}{\varepsilon}\ln\frac{1}{\varepsilon}$. The logarithmic term in this expression was attributed to the graphical complexity (i.e., the decoding complexity per iteration), and the number of iterations was conjectured to scale like $\frac{1}{\varepsilon}$. There is one exception: For the BEC, the complexity under the iterative message-passing decoding algorithm behaves like $\ln\frac{1}{\varepsilon}$ (see [51], [81], [80] and [94]). This is true since the absolute reliability provided by the BEC allows every edge in the graph to be used only once during the iterative decoding. Hence, for the BEC, the number of iterations performed by the decoder serves mainly to measure the delay in the decoding process, while the decoding complexity is closely related to the complexity of the Tanner graph which is chosen to represent the code. The graphical complexity required for LDPC and systematic irregular repeat-accumulate (IRA) code ensembles to achieve a fraction $1-\varepsilon$ of the capacity of a BEC under iterative decoding was studied in [81] and [80]. It was shown in these papers that the graphical complexity of these ensembles must scale at least like $\ln\frac{1}{\varepsilon}$; moreover, some explicit constructions were shown to approach the channel capacity with such a scaling of the graphical complexity. An additional degree of freedom which is obtained by introducing state nodes in the graph (e.g., punctured bits) was exploited in [65] and [64] to construct capacity-achieving ensembles of graph-based codes which achieve an improved tradeoff between complexity and achievable rates. Surprisingly, these capacity-achieving ensembles under iterative decoding were demonstrated to maintain a *bounded graphical complexity* regardless of the erasure probability of the BEC. A similar result of a bounded graphical complexity for capacity-achieving ensembles over the BEC was also obtained in [38].

This paper provides simple lower bounds on the number of iterations which is required for successful message-passing decoding of graph-based code ensembles. The transmission of these ensembles is assumed to take place over the BEC, and the bounds refer to the asymptotic case where the block length tends to infinity. The simplicity of the bounds derived in this paper stems from the fact that they are easily evaluated and are expressed in terms of some basic parameters of the considered

ensemble; these include the fraction of degree-2 variable nodes, the target bit erasure probability and the gap between the channel capacity and the design rate of the ensemble. The bounds derived in this paper demonstrate that the number of iterations which is required for successful message-passing decoding scales at least like the inverse of the gap (in rate) to capacity, provided that the fraction of degree-2 variable nodes of these turbo-like ensembles does not vanish (hence, the number of iterations becomes unbounded as the gap to capacity vanishes). The behavior of these lower bounds matches well with the experimental results and the conjectures on the number of iterations and complexity, as provided by Khandekar and McEliece (see [43, 42, 56]). Note that lower bounds on the number of iterations in terms of the target bit erasure probability can be alternatively viewed as lower bounds on the achievable bit erasure probability as a function of the number of iterations performed by the decoder. As a result of this, the simple bounds derived in this paper provide some insight on the design of stopping criteria for iteratively decoded ensembles over the BEC (for other stopping criteria see, e.g., [7, 90]).

This paper is structured as follows: Section 4.2 presents some preliminary background, definitions and notation, Section 4.3 introduces the main results of this paper and discusses some of their implications, the proofs of these statements and some further discussions are provided in Section 4.4. Finally, Section 4.5 summarizes this paper. Proofs of some technical statements are relegated to the appendices.

4.2 Preliminaries

This section provides preliminary background and introduces notation for the rest of this paper.

4.2.1 Graphical Complexity of Codes Defined on Graphs

As noted in Section 4.1, the decoding complexity of a graph-based code under iterative message-passing decoding is closely related to its graphical complexity, which we now define formally.

Definition 4.1 [**Graphical Complexity**] Let $\mathcal{C}$ be a binary linear block code of length n and rate R, and let $\mathcal{G}$ be an arbitrary representation of $\mathcal{C}$ by a Tanner graph. Denote the number of edges in $\mathcal{G}$ by E. The graphical complexity of $\mathcal{G}$ is defined as the number of edges in $\mathcal{G}$ per information bit of the code $\mathcal{C}$, i.e., $\Delta(\mathcal{G}) \triangleq \frac{E}{nR}$.

Note that the graphical complexity depends on the specific Tanner graph which is used to represent the code. An analysis of the graphical complexity for some families of graph-based codes is provided in [38, 65, 64, 81, 80].

4.2.2 Accumulate-Repeat-Accumulate Codes

Accumulate-repeat-accumulate (ARA) codes form an attractive coding scheme of turbo-like codes due to the simplicity of their encoding and decoding (where both scale linearly with the block length), and due to their remarkable performance under iterative decoding [4]. By some suitable constructions of puncturing patterns, ARA codes with small maximal node degree are presented in [4]; these codes perform very well even for short to moderate block lengths, and they suggest flexibility in the design of efficient rate-compatible codes operating on the same ARA decoder.

Ensembles of irregular and systematic ARA codes, which asymptotically achieve the capacity of the BEC with bounded graphical complexity, are presented in [64]. This bounded complexity result stays in contrast to LDPC ensembles, which have been shown to require unbounded graphical complexity in order to approach channel capacity, even under maximum-likelihood decoding (see [81]). In this section, we present ensembles of irregular and systematic ARA codes, and give a short overview of their encoding and decoding algorithms; this overview is required for the later discussion. The material contained in this section is taken from [64, Section II], and is introduced here briefly in order to make the paper self-contained.

From an encoding point of view, ARA codes are viewed as interleaved and serially concatenated codes. The encoding of ARA codes is done as follows: first, the information bits are accumulated (i.e., differentially encoded), and then the bits are repeated a varying number of times (by an irregular repetition code) and interleaved. The interleaved bits are partitioned into disjoint sets (whose size is not fixed in general), and the parity of each set of bits is computed (i.e., the bits are passed through an irregular single parity-check (SPC) code). Finally, the bits are accumulated a second time. A codeword of systematic ARA codes is composed of the information bits and the parity bits at the output of the second accumulator.

Since the iterative decoding algorithm of ARA codes is performed on the appropriate Tanner graph (see Figure 4.1), this leads one to view them as sparse-graph codes from a decoding point of view.

Following the notation in [64], we refer to the three layers of bit nodes in the Tanner graphs as 'systematic bits' which form the systematic part of the codeword, 'punctured bits' which correspond to the output of the first accumulator and are not a part of the transmitted codeword, and 'code bits' which correspond to the output of the second accumulator and form the parity-bits of the codeword (see Figure 4.1). Denoting the block length of the code by n and its dimension by k, each codeword is composed of k systematic bits and $n-k$ code bits. The two layers of check nodes are referred to as 'parity-check 1' nodes and 'parity-check 2' nodes, which correspond

to the first and the second accumulators of the encoder, respectively. An ensemble of irregular ARA codes is defined by the block length n and the degree distributions of the 'punctured bit' and 'parity-check 2' nodes. Following the notation in [64], the degree distribution of the 'punctured bit' nodes is given by the power series

$$L(x) \triangleq \sum_{i=1}^{\infty} L_i x^i \tag{4.1}$$

where L_i designates the fraction of 'punctured bit' nodes whose degree is i. Similarly, the degree distribution of the 'parity-check 2' nodes is given by

$$R(x) \triangleq \sum_{i=1}^{\infty} R_i x^i \tag{4.2}$$

where R_i designates the fraction of these nodes whose degree is i. In both cases, degree of a node only refers to edges connecting the 'punctured bit' and the 'parity-check 2' layers, without the extra two edges which are connected to each of the 'punctured bit' nodes and 'parity-check 2' nodes from the accumulators (see Figure 4.1). Considering the distributions from the edge perspective, we let

$$\lambda(x) \triangleq \sum_{i=1}^{\infty} \lambda_i x^{i-1}, \quad \rho(x) \triangleq \sum_{i=1}^{\infty} \rho_i x^{i-1} \tag{4.3}$$

designate the degree distributions from the edge perspective; here, λ_i (ρ_i) designates the fraction of edges connecting 'punctured bit' nodes to 'parity-check 2' nodes which are adjacent to 'punctured bit' ('parity-check 2') nodes of degree i. The design rate of a systematic ARA ensemble is given by $R = \frac{a_{\mathrm{R}}}{a_{\mathrm{L}}+a_{\mathrm{R}}}$ where

$$a_{\mathrm{L}} \triangleq \sum_i i L_i = L'(1) = \frac{1}{\int_0^1 \lambda(t)\mathrm{d}t}, \qquad a_{\mathrm{R}} \triangleq \sum_i i R_i = R'(1) = \frac{1}{\int_0^1 \rho(t)\mathrm{d}t} \tag{4.4}$$

designate the average degrees of the 'punctured bit' and 'parity-check 2' nodes, respectively.

Iterative decoding of ARA codes is performed by passing messages on the edges of the Tanner graph in a layer-by-layer approach. Each decoding iteration starts with messages for the 'systematic bit' nodes to the 'parity-check 1' nodes, the latter nodes then use this information to calculate new messages to the 'punctured bit' nodes and so the information passes through layers down the graph and back up until the iteration ends with messages from the 'punctured bit' nodes to the 'parity-check 1' nodes. The final phase of messages from the 'parity-check 1' nodes to the 'systematic bit' nodes is omitted since the latter nodes are of degree one and so the

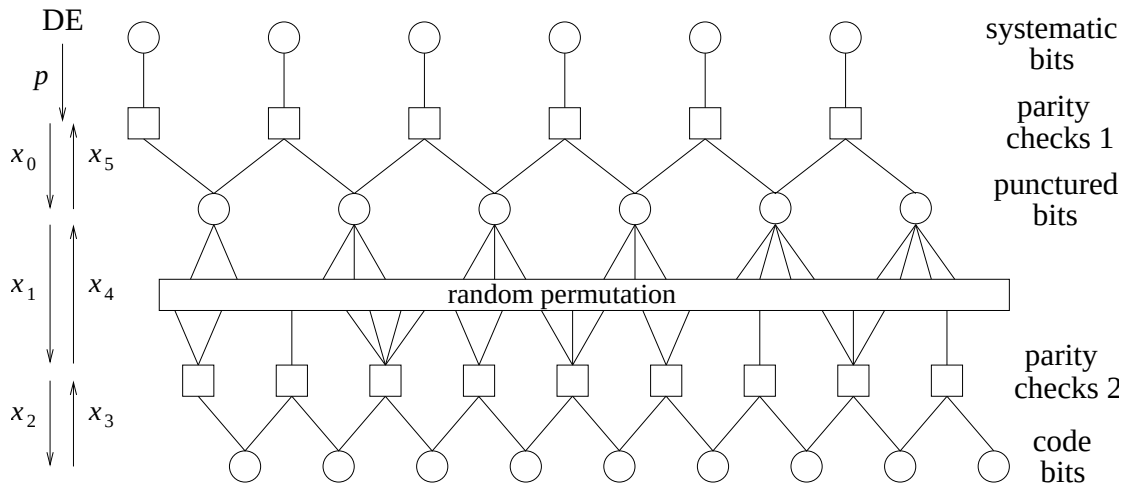

Figure 4.1: Tanner graph of an irregular and systematic accumulate-repeat-accumulate code. This figure is reproduced from [64].

outgoing message is not changed by incoming information. Assume that the code is transmitted over a BEC with erasure probability p. Since the systematic bits receive input from the channel, the probability of erasure in messages from the 'systematic bit' nodes to the 'parity-check 1' nodes is equal to p throughout the decoding process. For other messages, we denote by $x_i^{(l)}$ where $i = 0, 1, \ldots, 5$ the probability of erasure of the different message types at decoding iteration number l (where we start counting at zero). The variable $x_0^{(l)}$ corresponds to the probability of erasure in message from the 'parity-check 1' nodes to the 'punctured bit' nodes, $x_1^{(l)}$ tracks the erasure probability of messages from the 'punctured bit' nodes to the 'parity-check 2' nodes and so on. The density evolution (DE) equations for the decoder based on the Tanner graph in Figure 4.1 are given in [64], and we repeat them here:

$$\begin{aligned} x_0^{(l)} &= 1 - \left(1 - x_5^{(l-1)}\right)(1-p) \\ x_1^{(l)} &= \left(x_0^{(l)}\right)^2 \lambda\left(x_4^{(l-1)}\right) \\ x_2^{(l)} &= 1 - R\left(1 - x_1^{(l)}\right)\left(1 - x_3^{(l-1)}\right) \qquad l = 1, 2, \ldots \\ x_3^{(l)} &= p\, x_2^{(l)} \\ x_4^{(l)} &= 1 - \left(1 - x_3^{(l)}\right)^2 \rho\left(1 - x_1^{(l)}\right) \\ x_5^{(l)} &= x_0^{(l)}\, L\left(x_4^{(l)}\right). \end{aligned} \tag{4.5}$$

The stability condition for systematic ARA ensembles is derived in [64, Section II.D] and states that the fixed point $x_i^{(l)} = 0$ of the iterative decoding algorithm is stable

if and only if

$$p^2 \, \lambda_2 \, \left(\rho'(1) + \frac{2pR'(1)}{1-p} \right) \leq 1 \, . \tag{4.6}$$

4.2.3 Big-O notation

The terms O, Ω and Θ are widely used in computer science to describe asymptotic relationships between functions (for formal definitions see e.g., [1]). In our context, we refer to the gap (in rate) to capacity, denoted by ε, and discuss in particular the case where $0 \leq \varepsilon \ll 1$ (i.e., sequences of capacity-approaching ensembles). Accordingly, we define

- $f(\varepsilon) = O\big(g(\varepsilon)\big)$ means that there are positive constants c and δ, such that $0 \leq f(\varepsilon) \leq c \, g(\varepsilon)$ for all $0 \leq \varepsilon \leq \delta$.
- $f(\varepsilon) = \Omega\big(g(\varepsilon)\big)$ means that there are positive constants c and δ, such that $0 \leq c \, g(\varepsilon) \leq f(\varepsilon)$ for all $0 \leq \varepsilon \leq \delta$.
- $f(\varepsilon) = \Theta\big(g(\varepsilon)\big)$ means that there are positive constants c_1, c_2 and δ, such that $0 \leq c_1 \, g(\varepsilon) \leq f(\varepsilon) \leq c_2 \, g(\varepsilon)$ for all $0 \leq \varepsilon \leq \delta$.

Note that for all the above definitions, the values of c, c_1, c_2 and δ must be fixed for the function f and should not depend on ε.

4.3 Main Results

In this section, we present lower bounds on the required number of iterations used by a message-passing decoder for code ensembles defined on graphs. The communication is assumed to take place over a BEC, and we consider the asymptotic case where the block length of these code ensembles tends to infinity.

Definition 4.2 Let $\big\{\mathcal{C}_m\big\}_{m \in \mathbb{N}}$ be a sequence of code ensembles. Assume a common block length (n_m) of the codes in $\mathcal{C}_m$ which tends to infinity as m grows. Let the transmission of this sequence take place over a BEC with capacity C. The sequence $\big\{\mathcal{C}_m\big\}$ is said to *achieve a fraction* $1 - \varepsilon$ *of the channel capacity under some given decoding algorithm* if the asymptotic rate of the codes in $\mathcal{C}_m$ satisfies $R \geq (1-\varepsilon)C$ and the achievable bit erasure probability under the considered algorithm vanishes as m becomes large.

In the continuation, we consider a standard message-passing decoder for the BEC, and address the number of iterations which is required in terms of the achievable fraction of the channel capacity under this decoding algorithm.

Theorem 4.1 [Lower bound on the number of iterations for LDPC ensembles transmitted over the BEC] Let $\big\{(n_m, \lambda, \rho)\big\}_{m\in\mathbb{N}}$ be a sequence of LDPC ensembles whose transmission takes place over a BEC with erasure probability p. Assume that this sequence achieves a fraction $1-\varepsilon$ of the channel capacity under message-passing decoding. Let $L_2 = L_2(\varepsilon)$ be the fraction of variable nodes of degree 2 for this sequence. In the asymptotic case where the block length tends to infinity, let $l = l(\varepsilon, p, P_{\mathrm{b}})$ denote the number of iterations which is required to achieve an average bit erasure probability P_{b} over the ensemble. Under the mild condition that $P_{\mathrm{b}} < p\, L_2(\varepsilon)$, the required number of iterations satisfies the lower bound

$$l(\varepsilon, p, P_{\mathrm{b}}) \geq \frac{2}{1-p}\left(\sqrt{p\, L_2(\varepsilon)} - \sqrt{P_{\mathrm{b}}}\right)^2 \frac{1}{\varepsilon}\,. \tag{4.7}$$

Corollary 4.1 Under the assumptions of Theorem 4.1, if the fraction of degree-2 variable nodes stays strictly positive as the gap (in rate) to capacity vanishes, i.e., if

$$\lim_{\varepsilon\to 0} L_2(\varepsilon) > 0$$

then the number of iterations which is required in order to achieve an average bit erasure probability $P_{\mathrm{b}} < p\, L_2(\varepsilon)$ under iterative message-passing decoding scales at least like the inverse of this gap to capacity, i.e.,

$$l(\varepsilon, p, P_{\mathrm{b}}) = \Omega\left(\frac{1}{\varepsilon}\right).$$

Discussion 4.1 [Effect of messages' scheduling on the number of iterations] The lower bound on the number of iterations as provided in Theorem 4.1 refers to the *flooding schedule* where in each iteration, all the variable nodes and subsequently all the parity-check nodes send messages to their neighbors. Though it is the commonly used scheduling used by iterative message-passing decoding algorithms, an alternative scheduling of the messages may provide a faster convergence rate for the iterative decoder. As an example, [92] considers the convergence rate of a *serial scheduling* where instead of sending all the messages from the variable nodes to parity-check nodes and then all the messages from check nodes to variable nodes, as done in the flooding schedule, these two phases are interleaved. Based on the density evolution analysis which applies to the asymptotic case of an infinite block length, it is demonstrated in [92] that under some assumptions, the required number of iterations for LDPC decoding over the BEC with serial scheduling is reduced by a factor of two (as

compared to the flooding scheduling). It is noted that the main result of Theorem 4.1 is the introduction of a rigorous and simple lower bound on the number of iterations for LDPC ensembles which scales like the reciprocal of the gap between the channel capacity and the design rate of the ensemble. Though such a scaling of this bound is proved for the commonly used approach of flooding scheduling, it is likely to hold also for other efficient approaches of scheduling. It is also noted that this asymptotic scaling of the lower bound on the number of iterations supports the conjecture of Khandekar and McEliece [43].

Discussion 4.2 [On the dependence of the bounds on the fraction of degree-2 variable nodes] The lower bound on the number of iterations in Theorem 4.1 becomes trivial when the fraction of variable nodes of degree 2 vanishes. Let us focus our attention on sequences of ensembles which approach the channel capacity under iterative message-passing decoding (i.e., $\varepsilon \to 0$). For the BEC, several such sequences have been constructed (see e.g. [51, 94]). Asymptotically, as the gap to capacity vanishes, all of these sequences known to date satisfy the stability condition with equality; this property is known as the flatness condition [94]. In [76, Lemma 5], the asymptotic fraction of degree 2 variable nodes for capacity-approaching sequences of LDPC ensembles over the BEC is calculated. This lemma states that for such sequences which satisfy the following two conditions as the gap to capacity vanishes:

- The stability condition is satisfied with equality (i.e., the flatness condition holds)
- The limit of the ratio between the standard deviation and the expectation of the right degree exists and is finite

then the asymptotic fraction of degree–2 variable nodes does not vanish. In fact, for various sequences of capacity approaching LDPC ensembles known to date (see [51, 63, 94]), the ratio between the standard deviation and the expectation of the right degree-distribution tends to zero; in this case, [76, Lemma 5] implies that the fraction of degree-2 variable nodes tends to $\frac{1}{2}$ irrespectively of the erasure probability of the BEC, as can be verified directly for these code ensembles.

Discussion 4.3 [Concentration of the lower bound] Theorem 4.1 applies to the required number of iterations for achieving an average bit erasure probability P_{b} where this average is taken over the LDPC ensemble whose block length tends to infinity. Although we consider an expectation over the LDPC ensemble, note that l is deterministic as it is the smallest integer for which the average bit erasure probability

does not exceed a fixed value. As shown in the proof (see Section 4.4), the derivation of this lower bound relies on the density evolution technique which addresses the average performance of the ensemble. Based on concentration inequalities, it is proved that the performance of individual codes from the ensemble concentrates around the average performance over the ensemble as we let the block length tend to infinity [74, Appendix C]. In light of this concentration result and the use of density evolution in Section 4.4 (which applies to the case of an infinite block length), it follows that the lower bound on the number of iterations in Theorem 4.1 is valid with probability 1 for individual codes from the ensemble. This also holds for the ensembles of codes defined on graphs considered in Theorems 4.2 and 4.3.

Discussion 4.4 [On the number of required iterations for showing a mild improvement in the erasure probability during the iterative process] Note that for capacity-approaching LDPC ensembles, the lower bound on the number of iterations tells us that even for successfully starting the iteration process and reducing the bit erasure probability by a factor which is below the fraction of degree-2 variable nodes, the required number of iterations already scales like $\frac{1}{\varepsilon}$. This is also the behavior of the lower bound on the number of iterations even when the bit erasure probability should be made arbitrarily small; this lower bound therefore indicates that for capacity-approaching LDPC ensembles, a significant number of the iterations is performed for the starting process of the iterative decoding where the bit erasure probability is merely reduced by a factor of $\frac{1}{2}$ as compared to the erasure probability of the channel (see Discussion 4.2 as a justification for the one-half factor). This conclusion is also well interpreted by the area theorem and the asymptotic behavior of the two EXIT curves (for the variable nodes and the parity-check nodes) in the limit where $\varepsilon \to 0$; as the gap to capacity vanishes, both curves tend to be a step function jumping from 0 to 1 at the origin, so the iterations progress very slowly at the initial stages of the decoding process.

In the asymptotic case where we let the block length tend to infinity and the transmission takes place over the BEC, suitable constructions of capacity-achieving systematic ARA ensembles enable a fundamentally improved tradeoff between their graphical complexity and their achievable gap (in rate) to capacity under iterative decoding (see [64]). The graphical complexity of these systematic ARA ensembles remains bounded (and quite small) as the gap to capacity for these ensembles vanishes under iterative decoding; this stays in contrast to un-punctured LDPC code ensembles [81] and *systematic* irregular repeat-accumulate (IRA) ensembles [80] whose graphical complexity necessarily becomes unbounded as the gap to capacity vanishes (see [64,

Table I]). This observation raises the question whether the number of iterations which is required to achieve a desired bit erasure probability under iterative decoding, can be reduced by using systematic ARA ensembles. The following theorem provides a lower bound on the number of iterations required to achieve a desired bit erasure probability under message-passing decoding; it shows that similarly to the parallel result for LDPC ensembles (see Theorem 4.1), the required number of iterations for systematic ARA codes scales at least like the inverse of the gap to capacity.

Theorem 4.2 [Lower bound on the number of iterations for systematic ARA ensembles transmitted over the BEC] Let $\big\{(n_m, \lambda, \rho)\big\}_{m\in\mathbb{N}}$ be a sequence of systematic ARA ensembles whose transmission takes place over a BEC with erasure probability p. Assume that this sequence achieves a fraction $1-\varepsilon$ of the channel capacity under message-passing decoding. Let $L_2 = L_2(\varepsilon)$ be the fraction of 'punctured bit' nodes of degree 2 for this sequence (where the two edges related to the accumulator are not taken into account). In the asymptotic case where the block length tends to infinity, let $l = l(\varepsilon, p, P_{\mathrm{b}})$ designate the required number of iterations to achieve an average bit erasure probability P_{b} of the systematic bits. Under the mild condition that $1 - \sqrt{1 - \frac{P_{\mathrm{b}}}{p}} < p\, L_2(\varepsilon)$, the number of iterations satisfies the lower bound

$$l(\varepsilon, p, P_{\mathrm{b}}) \geq 2p(1-\varepsilon) \left(\sqrt{p\, L_2(\varepsilon)} - \sqrt{1 - \sqrt{1 - \frac{P_{\mathrm{b}}}{p}}} \right)^2 \frac{1}{\varepsilon}\,. \tag{4.8}$$

As noted in Section 4.2.2, systematic ARA codes can be viewed as serially concatenated codes where the systematic bits are associated with the outer code. These codes can be therefore decoded iteratively by using a turbo-like decoder for interleaved and serially concatenated codes. The following proposition states that the lower bound on the number of iterations in Theorem 4.2 is also valid for such an iterative decoder.

Proposition 4.1 [Lower bound on the number of iterations for systematic ARA codes under turbo-like decoding] Under the assumptions and notation of Theorem 4.2, the lower bound on the number of iterations in (4.8) is valid also when the decoding is performed by a turbo-like decoder for uniformly interleaved and serially concatenated codes.

The reader is referred to Appendix 4.A for a detailed proof. The following theorem which refers to irregular repeat-accumulate (IRA) ensembles is proved in a conceptually similar way to the proof of Theorem 4.2.

Theorem 4.3 [Lower bound on the number of iterations for IRA ensembles transmitted over the BEC] Let $\big\{(n_m, \lambda, \rho)\big\}_{m\in\mathbb{N}}$ be a sequence of (systematic or non-systematic) IRA ensembles whose transmission takes place over a BEC with erasure probability p. Assume that this sequence achieves a fraction $1-\varepsilon$ of the channel capacity under message-passing decoding. Let $L_2 = L_2(\varepsilon)$ be the fraction of 'information bit' nodes of degree 2 for this sequence. In the asymptotic case where the block length tends to infinity, let $l = l(\varepsilon, p, P_{\rm b})$ designate the required number of iterations to achieve an average bit erasure probability $P_{\rm b}$ of the information bits. For systematic codes, if $P_{\rm b} < p\, L_2(\varepsilon)$, then the number of iterations satisfies the lower bound

$$l(\varepsilon, p, P_{\rm b}) \geq 2(1-\varepsilon)\left(\sqrt{p\, L_2(\varepsilon)} - \sqrt{P_{\rm b}}\right)^2 \frac{1}{\varepsilon}. \tag{4.9}$$

For non-systematic codes, if $P_{\rm b} < L_2(\varepsilon)$, then

$$l(\varepsilon, p, P_{\rm b}) \geq 2(1-\varepsilon)\left(\sqrt{L_2(\varepsilon)} - \sqrt{P_{\rm b}}\right)^2 \frac{1}{\varepsilon}. \tag{4.10}$$

4.4 Derivation of the Bounds on the Number of Iterations

4.4.1 Proof of Theorem 4.1

Let $\big\{x^{(l)}\big\}_{l\in\mathbb{N}}$ designate the expected fraction of erasures in messages from the variable nodes to the check nodes at the l'th iteration of the message-passing decoding algorithm (where we start counting at $l=0$). From density evolution, in the asymptotic case where the block length tends to infinity, $x^{(l)}$ is given by the recursive equation

$$x^{(l+1)} = p\,\lambda\big(1-\rho\big(1-x^{(l)}\big)\big)\,, \quad l\in\mathbb{N} \tag{4.11}$$

with the initial condition

$$x^{(0)} = p \tag{4.12}$$

where p designates the erasure probability of the BEC. Considering a sequence of $\{(n_m, \lambda, \rho)\}$ LDPC ensembles where we let the block length n_m tend to infinity, the average bit erasure probability after the l'th iteration is given by

$$P_{\rm b}^{(l)} = p\, L\big(1-\rho(1-x^{(l)})\big) \tag{4.13}$$

where L designates the common left degree distribution of the ensembles from the node perspective. Since the function $f(x) = p\,\lambda\big(1-\rho(1-x)\big)$ is monotonically increasing, Eqs. (4.11)–(4.13) imply that an average bit erasure probability of $P_{\rm b}$ is attainable under iterative message-passing decoding if and only if

$$p\,\lambda\big(1-\rho(1-x)\big) < x\,, \quad \forall x\in(x^*, p] \tag{4.14}$$

where x^* is the unique solution of

$$P_{\mathrm{b}} = p\, L\big(1-\rho(1-x^*)\big)\,.$$

Let us define the functions

$$c(x) \triangleq 1-\rho(1-x), \qquad v(x) = \begin{cases} \lambda^{-1}\left(\frac{x}{p}\right) & 0 \le x \le p \\ 1 & p < x \le 1 \end{cases}\,. \tag{4.15}$$

From the condition in (4.14), an average bit erasure probability of P_{b} is attained if and only if $c(x) < v(x)$ for all $x \in (x^*, p]$. Since we assume that vanishing bit erasure probability is achievable under message-passing decoding, it follows that $c(x) < v(x)$ for all $x \in (0, p]$. Figure 4.2 shows a plot of the functions $c(x)$ and $v(x)$ for an ensemble of LDPC codes which achieves vanishing bit erasure probability under iterative decoding as the block length tends to infinity. The horizontal and vertical lines, labeled $\{h_l\}_{l\in\mathbb{N}}$ and $\{v_l\}_{l\in\mathbb{N}}$, respectively, are used to track the expected fraction of erased messages from the variable nodes to the parity-check nodes at each iteration of the message-passing decoding algorithm. From (4.11) and (4.12), the expected fraction of erased left to right messages in the l'th decoding iteration (where we start counting at zero) is equal to the x value at the left tip of the horizontal line h_l. The right-angled triangles shaded in gray will be used later in the proof.

The first step in the proof of Theorem 4.1 is calculating the area bounded by the curves $c(x)$ and $v(x)$. This is done in the following lemma which is based on the area theorem for the BEC [9].

Lemma 4.1

$$\int_0^1 \big(v(x)-c(x)\big)\mathrm{d}x = \frac{C-R}{a_{\mathrm{L}}} \tag{4.16}$$

where $C = 1-p$ is the capacity of the BEC, R is the design rate of the ensemble, and a_{L} is the average left degree of the ensemble.

Proof: The definitions of the functions v and c in (4.15) imply that

$$\begin{aligned}
\int_0^1 \big(v(x)-c(x)\big)\mathrm{d}x &= \int_0^p \lambda^{-1}\left(\frac{x}{p}\right)\mathrm{d}x + \int_p^1 1\,\mathrm{d}x - \int_0^1 c(x)\mathrm{d}x \\
&= p\int_0^1 \lambda^{-1}(s)\mathrm{d}s + 1 - p - \int_0^1 \big(1-\rho(1-x)\big)\mathrm{d}x \\
&\overset{(a)}{=} p\left(1-\int_0^1 \lambda(x)\mathrm{d}x\right) + 1 - p - 1 + \int_0^1 \rho(x)\mathrm{d}x \\
&= \int_0^1 \rho(x)\mathrm{d}x - p\int_0^1 \lambda(x)\mathrm{d}x \\
&= \underbrace{\int_0^1 \lambda(x)\mathrm{d}x}_{\frac{1}{a_{\mathrm{L}}}} \left(\underbrace{\frac{\int_0^1 \rho(x)\mathrm{d}x}{\int_0^1 \lambda(x)\mathrm{d}x}}_{1-R} - \underbrace{p}_{1-C}\right) \\
&= \frac{C-R}{a_{\mathrm{L}}}
\end{aligned}$$

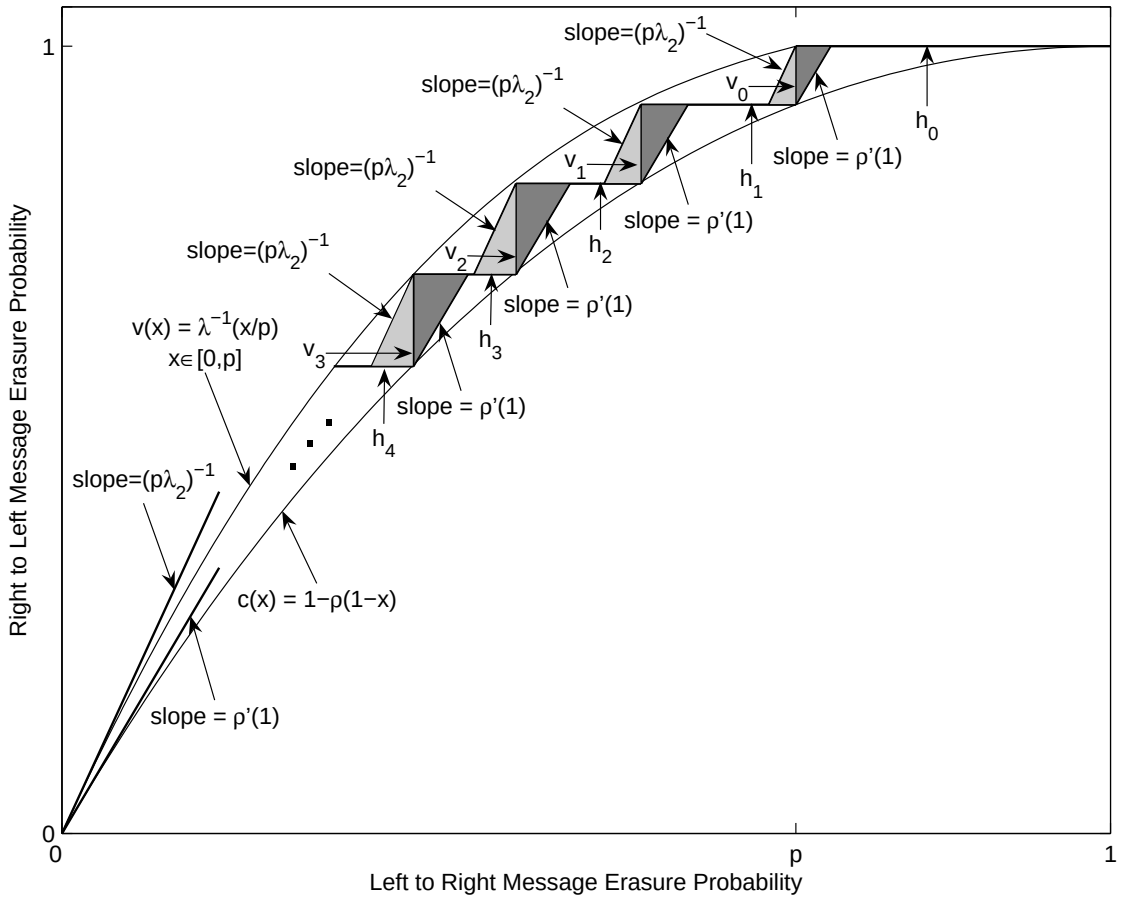

Figure 4.2: Plot of the functions $c(x)$ and $v(x)$ for an ensemble of LDPC codes which achieves vanishing bit erasure probability under iterative message-passing decoding when communicated over a BEC whose erasure probability is equal to p. The horizontal and vertical lines track the evolution of the expected fraction of erasure messages from the variable nodes to the check nodes at each iteration of the message-passing decoding algorithm.

where (a) follows by substituting $x = \lambda(s)$ and applying integration by parts. ■

Let us consider the two sets of right-angled triangles shown in two shades of gray in Figure 4.2. The set of triangles which are shaded in dark gray are defined so that one of the legs of triangle number i (counting from right to left and starting at zero) is the vertical line v_i, and the slope of the hypotenuse is equal to $c'(0) = \rho'(1)$. Since $c(x)$ is concave for all $x \in [0, 1]$, these triangles are guaranteed to be above the curve of the function c. Since the slope of the hypotenuse is $\rho'(1)$, the area of the i'th triangle in this set is

$$A_i = \frac{1}{2} \, |v_i| \, \left(\frac{|v_i|}{\rho'(1)} \right) = \frac{|v_i|^2}{2\rho'(1)} \tag{4.17}$$

where $|v_i|$ is the length of v_i. We now turn to consider the second set of triangles, which are shaded in light gray. Note that the function $\lambda(x)$ is monotonically increasing and convex in $[0,1]$ and also that $\lambda(0) = 0$ and $\lambda(1) = 1$. This implies that λ^{-1} is concave in $[0,1]$ and therefore $v(x)$ is concave in $[0,p]$. The triangles shaded in light gray are defined so that one of the legs of triangle number i (again, counting from the right and starting at zero) is the vertical line v_i and the slope of the hypotenuse is given by

$$v'(0) = \frac{1}{p}\left(\lambda^{-1}\right)'(0) = \frac{1}{p\lambda'(0)} = \frac{1}{p\lambda_2}$$

where the second equality follows since $\lambda(0) = 0$. The concavity of $v(x)$ in $[0,p]$ guarantees that these triangles are below the curve of the of function v. The area of the i'th triangle in this second set of triangles is given by

$$B_i = \frac{1}{2}\,|v_i|\;\left(|v_i|\,p\lambda_2\right) = \frac{p\lambda_2\,|v_i|^2}{2}\,. \tag{4.18}$$

Since $v(x)$ is monotonically increasing with x, the dark-shaded triangles lie below the curve of the function v. Similarly, the monotonicity of $c(x)$ implies that the light-shaded triangles are above the curve of the function c. Hence, both sets of triangles form a subset of the domain bounded by the curves of $c(x)$ and $v(x)$. By their definitions, the i'th dark triangle is on the right of v_i, and the i'th light triangle lies to the left of v_i; therefore, the triangles do not overlap. Combining (4.17), (4.18) and the fact that the triangles do not overlap, and applying Lemma 4.1, we get

$$\begin{aligned}
\frac{C-R}{a_{\mathrm{L}}} &= \int_0^1 \big(v(x)-c(x)\big)\mathrm{d}x \\
&\geq \sum_{i=0}^{\infty}(A_i+B_i) \\
&\geq \frac{1}{2}\left(\frac{1}{\rho'(1)}+p\lambda_2\right)\sum_{i=0}^{l-1}|v_i|^2
\end{aligned} \tag{4.19}$$

where l is an arbitrary natural number. Since we assume that the bit erasure probability vanishes under iterative message-passing decoding, the stability condition implies that

$$\frac{1}{\rho'(1)} \geq p\,\lambda_2\,. \tag{4.20}$$

Substituting (4.20) and $R = (1-\varepsilon)C$ in (4.19) gives

$$C\varepsilon \geq a_{\mathrm{L}}\,p\lambda_2\sum_{i=0}^{l-1}|v_i|^2. \tag{4.21}$$

The definition of h_l and v_l in Figure 4.2 implies that for an arbitrary iteration l

$$1 - \rho(1 - x^{(l)}) = c(x^{(l)}) = 1 - \sum_{i=0}^{l} |v_i| \,.$$

Substituting the last equality in (4.13) yields that the average bit erasure probability after iteration number $l-1$ can be expressed as

$$P_{\mathrm{b}}^{(l-1)} = p \; L \left(1 - \sum_{i=0}^{l-1} |v_i| \right) . \tag{4.22}$$

Let l designate the number of iterations required to achieve an average bit erasure probability P_{b} over the ensemble (where we let the block length tend to infinity), i.e., l is the smallest integer which satisfies $P_{\mathrm{b}}^{(l-1)} \leq P_{\mathrm{b}}$ since we start counting at $l = 0$. Although we consider an expectation over the LDPC ensemble, note that l is deterministic as it is the smallest integer for which the average bit erasure probability does not exceed P_{b}. Since L is monotonically increasing, (4.22) provides a lower bound on $\sum_{i=0}^{l-1} |v_i|$ of the form

$$\sum_{i=0}^{l-1} |v_i| \geq 1 - L^{-1} \left(\frac{P_{\mathrm{b}}}{p} \right) . \tag{4.23}$$

From the Cauchy-Schwartz inequality, we get

$$\left(\sum_{i=0}^{l-1} |v_i| \right)^2 \leq \sum_{i=0}^{l-1} 1 \sum_{i=0}^{l-1} |v_i|^2 = l \sum_{i=0}^{l-1} |v_i|^2. \tag{4.24}$$

Combining the above inequality with (4.21) and (4.23) gives the inequality

$$C\varepsilon \geq \frac{a_{\mathrm{L}} \, p\lambda_2 \left(1 - L^{-1}\left(\frac{P_{\mathrm{b}}}{p} \right) \right)^2}{l}$$

which provides the following lower bound on the number of iterations l:

$$l \geq \frac{a_{\mathrm{L}} \, p\lambda_2 \left(1 - L^{-1}\left(\frac{P_{\mathrm{b}}}{p} \right) \right)^2}{(1-p)\varepsilon}. \tag{4.25}$$

To continue the proof, we derive a lower bound on $1 - L^{-1}(x)$ for $x \in (0, 1)$. Since the fraction of variable nodes of degree i is non-negative for all $i = 2, 3, \ldots$, we have

$$L(x) = \sum_i L_i x^i \geq L_2 x^2, \quad x \geq 0.$$

Substituting $t = L(x)$ gives

$$t \geq L_2 \cdot \left(L^{-1}(t) \right)^2, \quad \forall t \in (0, 1)$$

which is transformed into the following lower bound on $1 - L^{-1}(x)$:

$$1 - L^{-1}(x) \geq 1 - \sqrt{\frac{x}{L_2}}, \quad \forall x \in (0,1)\,. \tag{4.26}$$

Under the assumption $\frac{P_{\rm b}}{p} < L_2$, substituting (4.26) in (4.25) gives

$$\begin{aligned} l &\geq \frac{a_{\rm L}\, p\lambda_2 \left(\sqrt{L_2} - \sqrt{\frac{P_{\rm b}}{p}}\right)^2}{L_2\,(1-p)\varepsilon} \\ &= \frac{a_{\rm L}\,\lambda_2 \left(\sqrt{p\,L_2} - \sqrt{P_{\rm b}}\right)^2}{L_2\,(1-p)\varepsilon}\,. \end{aligned} \tag{4.27}$$

The lower bound in (4.7) is obtained by substituting the equality $L_2 = \frac{\lambda_2\, a_{\rm L}}{2}$ into (4.27).

Taking the limit where the average bit erasure probability tends to zero on both sides of (4.7) gives the following lower bound on the number of iterations:

$$l(\varepsilon, p, P_{\rm b} \to 0) \geq \frac{2p}{1-p}\frac{L_2(\varepsilon)}{\varepsilon}\,.$$

4.4.2 Proof of Theorem 4.2

We begin the proof by considering the expected fraction of erasure messages from the 'punctured bit' nodes to the 'parity-check 2' nodes (see Figure 4.1). The following lemma provides a lower bound on the expected fraction of erasures in the l'th decoding iteration in terms of this expected fraction at the preceding iteration.

Lemma 4.2 Let (n, λ, ρ) be an ensemble of systematic ARA codes whose transmission takes place over a BEC with erasure probability p. Then, in the limit where the block length tends to infinity, the expected fraction of erasure messages from the 'punctured bit' nodes to the 'parity-check 2' nodes at the l'th iteration satisfies

$$x_1^{(l)} \geq \widetilde{\lambda}\left(1 - \widetilde{\rho}\big(1 - x_1^{(l-1)}\big)\right), \quad l = 1, 2, \ldots \tag{4.28}$$

where the tilted degree distributions $\widetilde{\lambda}$ and $\widetilde{\rho}$ are given as follows (see [64]):

$$\widetilde{\lambda}(x) \triangleq \left(\frac{p}{1-(1-p)L(x)}\right)^2 \lambda(x) \tag{4.29}$$

$$\widetilde{\rho}(x) \triangleq \left(\frac{1-p}{1-pR(x)}\right)^2 \rho(x) \tag{4.30}$$

and L and R designate the degree distributions of the ARA ensemble from the node perspective.

Proof: See Appendix 4.B.1. ■

From Figure 4.1, it can be readily verified that the probabilities x_0 and x_1 for erasure messages at iteration no. zero are equal to 1, i.e.,

$$x_0^{(0)} = x_1^{(0)} = 1. \tag{4.31}$$

Let us look at the RHS of (4.28) as a function of x, and observe that it is monotonically increasing over the interval $[0, 1]$. Let us compare the performance of a systematic ARA ensemble whose degree distributions are (λ, ρ) with an LDPC ensemble whose degree distributions are given by $(\widetilde{\lambda}, \widetilde{\rho})$ (see (4.29) and (4.30)) under iterative message-passing decoding. Given the initial condition $x_1^{(0)} = 1$, the following conclusion is obtained by recursively applying Lemma 4.2: For any iteration, the erasure probability for messages delivered from 'punctured bit' nodes to 'parity-check 2' nodes of the ARA ensemble (see Figure 4.1) is lower bounded by the erasure probability of the left-to-right messages of the LDPC ensemble; this holds even if the a-priori information from the BEC is not used by the iterative decoder of the LDPC ensemble (note that the coefficient of $\widetilde{\lambda}$ in the RHS of (4.28) is equal to one). Note that unless the fraction of 'parity-check 2' nodes of degree 1 is strictly positive (i.e., $R_1 > 0$), the iterative decoding cannot be initiated for both ensembles (unless some the values of some 'punctured bits' of the systematic ARA ensemble are known, as in [64]). Hence, the comparison above between the ARA and LDPC ensembles is of interest under the assumption that $R_1 > 0$; this property is implied by the assumption of vanishing bit erasure probability for the systematic ARA ensemble under iterative message-passing decoding.

In [64, Section II.C.2], a technique called 'graph reduction' is introduced. This technique transforms the Tanner graph of a systematic ARA ensemble, transmitted over a BEC whose erasure probability is p, into a Tanner graph of an equivalent LDPC ensemble (where this equivalence holds in the asymptotic case where the block length tends to infinity). The variable and parity-check nodes of the equivalent LDPC code evolve from the 'punctured bit' and 'parity-check 2' nodes of the ARA ensemble, respectively, and their degree distributions (from the edge perspective) are given by $\widetilde{\lambda}$ and $\widetilde{\rho}$, respectively. It is also shown in [64] that $\widetilde{\lambda}$ and $\widetilde{\rho}$ are legitimate degree distribution functions, i.e., all the derivatives at zero are non-negative and $\widetilde{\lambda}(1) = \widetilde{\rho}(1) = 1$. As shown in [64, Eqs. (9)–(12)], the left and right degree distributions of the equivalent LDPC ensemble from the node perspective are given, respectively, by

$$\widetilde{L}(x) = \frac{\displaystyle\int_0^x \widetilde{\lambda}(t)\mathrm{d}t}{\displaystyle\int_0^1 \widetilde{\lambda}(t)\mathrm{d}t} = \frac{p\, L(x)}{1 - (1-p)L(x)} \tag{4.32}$$

and

$$\widetilde{R}(x) = \frac{\int_0^x \widetilde{\rho}(t)\mathrm{d}t}{\int_0^1 \widetilde{\rho}(t)\mathrm{d}t} = \frac{(1-p)\,R(x)}{1-pR(x)}\,. \tag{4.33}$$

Let $P_{\mathrm{b}}^{(l)}$ designate the average erasure probability of the systematic bits after the l'th decoding iteration (where we start counting at $l = 0$). For LDPC ensembles, a simple relationship between the erasure probability of the code bits and the erasure probability of the left-to-right messages at the l'th decoding iteration is given in (4.13). For systematic ARA ensembles, a similar, though less direct, relationship exists between the erasure probability of the systematic bits after the l'th decoding iteration and $x_1^{(l)}$; this relationship is presented in the following lemma.

Lemma 4.3 Let (n, λ, ρ) be an ensemble of systematic ARA codes whose transmission takes place over a BEC with erasure probability p. Then, in the asymptotic case where the block length tends to infinity, the average erasure probability of the systematic bits after the l'th decoding iteration, $P_{\mathrm{b}}^{(l)}$, satisfies the inequality

$$1 - \sqrt{1 - \frac{P_{\mathrm{b}}^{(l)}}{p}} \geq \widetilde{L}\left(1 - \widetilde{\rho}\left(1 - x_1^{(l)}\right)\right) \tag{4.34}$$

where $\widetilde{\rho}$ and $\widetilde{L}$ are defined in (4.30) and (4.32), respectively (similarly to their definitions in [64]).

Proof: See Appendix 4.B.2. ∎

Remark 4.1 We note that when $P_{\mathrm{b}}^{(l)}$ is very small, the LHS of (4.34) satisfies

$$1 - \sqrt{1 - \frac{P_{\mathrm{b}}^{(l)}}{p}} \approx \frac{P_{\mathrm{b}}^{(l)}}{2p}\,,$$

so (4.34) takes a similar form to (4.13) which refers to the erasure probability of LDPC ensembles.

Consider the number of iterations required for the message-passing decoder, operating on the Tanner graphs of the systematic ARA ensemble, to achieve a desired bit erasure probability P_{b}. Combining Lemmas 4.2 and 4.3, and the initial condition in (4.31), a lower bound on this number of iterations can be deduced. More explicitly, it is lower bounded by the number of iterations which is required to achieve a bit erasure probability of $1 - \sqrt{1 - \frac{P_{\mathrm{b}}}{p}}$ for the LDPC ensemble whose degree distributions are $(\widetilde{\lambda}, \widetilde{\rho})$ and where the erasure probability of the BEC is equal to 1. It is therefore

tempting to apply the lower bound on the number of iterations in Theorem 4.1, which refers to LDPC ensembles, as a lower bound on the number of iterations for the ARA ensemble. Unfortunately, the LDPC ensemble with the tilted pair of degree distributions $(\widetilde{\lambda}, \widetilde{\rho})$ is transmitted over a BEC whose erasure probability is 1, so the channel capacity is equal to zero and the multiplicative gap to capacity is meaningless. This prevents a direct use of Theorem 4.1; however, the continuation of the proof follows similar lines in the proof of Theorem 4.1.

Let x^* denote the unique solution in $[0, 1]$ of the equation

$$1 - \sqrt{1 - \frac{P_{\mathrm{b}}}{p}} = \widetilde{L}\big(1 - \widetilde{\rho}\,(1 - x^*)\big). \tag{4.35}$$

From (4.28), (4.31) and (4.34), a necessary condition for achieving a bit erasure probability P_{b} of the systematic bits is that

$$\widetilde{\lambda}\big(1 - \widetilde{\rho}(1 - x)\big) < x\,, \quad \forall x \in (x^*, 1]\,. \tag{4.36}$$

In the limit where the fixed point of the iterative decoding process is attained, the inequalities in (4.28), (4.31) and (4.34) are replaced by equalities; hence, (4.36) also forms a sufficient condition. Analogously to the case of LDPC ensembles, as in the proof of Theorem 4.1, we define the functions

$$\widetilde{c}(x) = 1 - \widetilde{\rho}(1 - x) \quad \text{and} \quad v(x) = \widetilde{\lambda}^{-1}(x)\,. \tag{4.37}$$

Due to the monotonicity of $\widetilde{\lambda}$ in $[0, 1]$, the necessary and sufficient condition for attaining an erasure probability P_{b} of the systematic bits in (4.36) can be rewritten as

$$\widetilde{c}(x) < \widetilde{v}(x)\,, \quad \forall x \in (x^*, 1]\,.$$

Since we assume that the sequence of ensembles asymptotically achieves vanishing bit erasure probability under message-passing decoding, it follows that

$$\widetilde{c}(x) < \widetilde{v}(x)\,, \quad \forall x \in (0, 1]\,.$$

The next step in the proof is calculating the area of the domain bounded by the curves $\widetilde{c}(x)$ and $\widetilde{v}(x)$. This is done in the following lemma which is analogous to Lemma 4.1.

Lemma 4.4

$$\int_0^1 \big(\widetilde{v}(x) - \widetilde{c}(x)\big)\mathrm{d}x = \frac{C - R}{(1 - R)\,a_{\mathrm{R}}} \tag{4.38}$$

where $\widetilde{v}$ and $\widetilde{c}$ are introduced in (4.37), $C = 1 - p$ is the capacity of the BEC, R is the design rate of the systematic ARA ensemble, and a_{R} is defined in (4.4) and it designates the average degree of the 'parity-check 2' nodes when the two edges related to the lower accumulator in Figure 4.1 are not taken into account.

Proof: The definitions of the functions $\widetilde{v}$ and $\widetilde{c}$ in (4.37) yield that

$$\begin{aligned}\int_0^1 \big(\widetilde{v}(x)-\widetilde{c}(x)\big)\mathrm{d}x &= \int_0^1 \widetilde{\lambda}^{-1}(x)\mathrm{d}x - 1 + \int_0^1 \widetilde{\rho}(1-x)\mathrm{d}x \\ &= \left(1-\int_0^1 \widetilde{\lambda}(x)\mathrm{d}x\right) - 1 + \int_0^1 \widetilde{\rho}(x)\mathrm{d}x \\ &= \int_0^1 \widetilde{\rho}(x)\mathrm{d}x - \int_0^1 \widetilde{\lambda}(x)\mathrm{d}x \end{aligned} \tag{4.39}$$

where the second equality is obtained via integration by parts (note that $\widetilde{\lambda}(0)=0$ and $\widetilde{\lambda}(1)=1$). From (4.32), we get

$$\int_0^1 \widetilde{\lambda}(x)\mathrm{d}x = \frac{1}{\widetilde{L}'(1)} = \frac{p}{L'(1)} = \frac{p}{a_{\mathrm{L}}} \tag{4.40}$$

(see also [64, Eq. (23)]) where a_{L} is defined in (4.4) and designates the average degree of the 'punctured bit' nodes in the Tanner graph (see Figure 4.1) when the two edges, related to the upper accumulator in Figure 4.1, are not taken into account. Similarly, (4.33) gives

$$\int_0^1 \widetilde{\rho}(x)\mathrm{d}x = \frac{1}{\widetilde{R}'(1)} = \frac{1-p}{R'(1)} = \frac{1-p}{a_{\mathrm{R}}} \tag{4.41}$$

(see also [64, Eq. (24)]). Substituting (4.40) and (4.41) into (4.39) gives

$$\begin{aligned}\int_0^1 \big(\widetilde{v}(x)-\widetilde{c}(x)\big)\mathrm{d}x &= \frac{1-p}{a_{\mathrm{R}}} - \frac{p}{a_{\mathrm{L}}} \\ &\overset{(a)}{=} \frac{1}{a_{\mathrm{R}}}\left[1 - p\underbrace{\left(\frac{a_{\mathrm{L}}+a_{\mathrm{R}}}{a_{\mathrm{L}}}\right)}_{\frac{1}{1-R}}\right] \\ &= \frac{1}{a_{\mathrm{R}}}\,\frac{1-R-p}{1-R} \\ &= \frac{C-R}{(1-R)\,a_{\mathrm{R}}} \end{aligned} \tag{4.42}$$

where (a) follows since the design rate of the systematic ARA ensemble is given by $R = \frac{a_{\mathrm{R}}}{a_{\mathrm{L}}+a_{\mathrm{R}}}$ (this equality follows directly from Figure 4.1). ■

To continue the proof, we consider a plot similar to the one in Figure 4.2 with the exception that $c(x)$ and $v(x)$ are replaced by $\widetilde{c}(x)$ and $\widetilde{v}(x)$, respectively. Note that in this case the horizontal line h_0 is reduced to the point $(1,1)$. Consider the two sets of gray-shaded right-angled triangles. The triangles shaded in dark gray are defined so that the height of triangle number i (counting from right to left and starting at zero) is the vertical line v_i and the slope of their hypotenuse is equal to $\widetilde{c}'(0) = \widetilde{\rho}'(1)$.

Since $\widetilde{c}(x)$ is concave, these triangles form a subset of the domain bounded by the curves $\widetilde{c}(x)$ and $\widetilde{v}(x)$. The area of the i'th triangle in this set is given by

$$A_i = \frac{1}{2}\,|v_i|\,\left(\frac{|v_i|}{\widetilde{\rho}'(1)}\right) = \frac{|v_i|^2}{2\,\widetilde{\rho}'(1)}$$

where $|v_i|$ is the length of v_i. The second set of right-angled triangles, which are shaded in light gray, are also defined so that the height of the i'th triangle (counting from right to left and starting at zero) is the vertical line v_i, but the triangle lies to the left of v_i and the slope of its hypotenuse is equal to

$$\widetilde{v}'(0) = \left(\widetilde{\lambda}^{-1}\right)'(0) = \frac{1}{\widetilde{\lambda}'(0)} = \frac{1}{p^2\lambda'(0)} = \frac{1}{p^2\lambda_2}$$

where the second equality follows since $\widetilde{\lambda}(0) = 0$ and the third equality follows from the definition of $\widetilde{\lambda}$ in (4.29). Since $\widetilde{\lambda}$ is monotonically increasing and convex over the interval $[0, 1]$ and it satisfies $\widetilde{\lambda}(0) = 0$ and $\widetilde{\lambda}(1) = 1$, then it follows that $v(x) = \widetilde{\lambda}^{-1}(x)$ is concave over this interval. Hence, the triangles shaded in light gray also form a subset of the domain bounded by the curves $c(x)$ and $v(x)$. The area of the i'th light-gray triangle is given by

$$B_i = \frac{1}{2}\,|v_i|\,\left(|v_i|\,p^2\lambda_2\right) = \frac{p^2\lambda_2\,|v_i|^2}{2}$$

Applying Lemma 4.4 and the fact that the triangles in both sets do not overlap, we get

$$\frac{C-R}{(1-R)\,a_{\mathrm{R}}} \geq \frac{1}{2}\left(\frac{1}{\widetilde{\rho}'(1)} + p^2\lambda_2\right)\sum_{i=0}^{l-1}|v_i|^2 \tag{4.43}$$

where l is an arbitrary natural number. Since the sequence of ensembles asymptotically achieves vanishing bit erasure probability under iterative message-passing decoding, the stability condition for systematic ARA codes (see (4.6) or equivalently [64, Eq. (14)]) implies that

$$p^2\lambda_2 \leq \frac{1}{\rho'(1) + \frac{2pR'(1)}{1-p}} = \frac{1}{\widetilde{\rho}'(1)} \tag{4.44}$$

where the last equality follows from (4.30). Substituting (4.44) in (4.43) gives

$$\frac{C-R}{(1-R)\,a_{\mathrm{R}}} \geq p^2\lambda_2\sum_{i=0}^{l-1}|v_i|^2\,. \tag{4.45}$$

Let $x^{(l)}$ denote the x value of the left tip of the horizontal line h_l. The value of $x^{(l)}$ satisfies the recursive equation

$$x^{(l+1)} = \widetilde{\lambda}\Big(1 - \widetilde{\rho}\big(1 - x^{(l)}\big)\Big), \quad \forall\, l \in \mathbb{N} \tag{4.46}$$

with $x^{(0)} = 1$. As was explained above (immediately following Lemma 4.2), from (4.28), (4.31), and the monotonicity of the function $f(x) = \widetilde{\lambda}\big(1 - \widetilde{\rho}(1-x)\big)$ over the interval $[0,1]$, we get that $x^{(l)} \leq x_1^{(l)}$ for $l \in \mathbb{N}$. The definition of h_l and v_l in Figure 4.2 implies that

$$1 - \widetilde{\rho}\big(1 - x^{(l)}\big) = \widetilde{c}\big(x^{(l)}\big) = 1 - \sum_{i=0}^{l} |v_i| \,. \tag{4.47}$$

Starting from (4.34) and applying the monotonicity of $\widetilde{L}$ and $\widetilde{\rho}$ gives

$$\begin{aligned} 1 - \sqrt{1 - \frac{P_{\mathrm{b}}^{(l-1)}}{p}} &\geq \widetilde{L}\Big(1 - \widetilde{\rho}\big(1 - x_1^{(l-1)}\big)\Big) \\ &\geq \widetilde{L}\left(1 - \widetilde{\rho}\big(1 - x^{(l-1)}\big)\right) \\ &= \widetilde{L}\left(1 - \sum_{i=0}^{l-1} |v_i|\right) \end{aligned}$$

where the last equality follows from (4.47). Since $\widetilde{L}$ is strictly monotonically increasing in $[0,1]$, then

$$\sum_{i=0}^{l-1} |v_i| \geq 1 - \widetilde{L}^{-1}\left(1 - \sqrt{1 - \frac{P_{\mathrm{b}}^{(l-1)}}{p}}\right). \tag{4.48}$$

Applying the Cauchy-Schwartz inequality (as in (4.24)) to the RHS of (4.45), we get

$$\begin{aligned} \frac{C-R}{(1-R)\,a_{\mathrm{R}}} &\geq p^2 \lambda_2 \sum_{i=0}^{l-1} |v_i|^2 \\ &\geq \frac{p^2 \lambda_2}{l} \left(\sum_{i=0}^{l-1} |v_i|\right)^2 \\ &\geq \frac{p^2 \lambda_2}{l} \left(1 - \widetilde{L}^{-1}\left(1 - \sqrt{1 - \frac{P_{\mathrm{b}}^{(l-1)}}{p}}\right)\right)^2 \end{aligned}$$

where the last inequality follows from (4.48). Since the design rate R is assumed to be a fraction $1 - \varepsilon$ of the capacity of the BEC, the above inequality gives

$$C\varepsilon \geq \frac{p^2 \lambda_2 \,(1-R)\, a_{\mathrm{R}} \left(1 - \widetilde{L}^{-1}\left(1 - \sqrt{1 - \frac{P_{\mathrm{b}}^{(l-1)}}{p}}\right)\right)^2}{l}$$

where l is an arbitrary natural number. Let l designate the number of iterations required to achieve an average bit erasure probability P_{b} of the systematic bits, i.e.,

l is the smallest integer which satisfies $P_{\mathrm{b}}^{(l-1)} \leq P_{\mathrm{b}}$ (since we start counting the iterations at $l = 0$). Note that l is deterministic since it refers to the smallest number of iterations required to achieve a desired average bit erasure probability over the ensemble. From the inequality above and the monotonicity of $\widetilde{L}$, we obtain that

$$C\varepsilon \geq \frac{p^2\lambda_2\,(1-R)\,a_{\mathrm{R}}\,\left(1-\widetilde{L}^{-1}\left(1-\sqrt{1-\frac{P_{\mathrm{b}}}{p}}\right)\right)^2}{l}$$

which provides a lower bound on the number of iterations of the form

$$\begin{aligned} l &\geq \frac{p^2\lambda_2(1-R)\,a_{\mathrm{R}}\,\left(1-\widetilde{L}^{-1}\left(1-\sqrt{1-\frac{P_{\mathrm{b}}}{p}}\right)\right)^2}{C\varepsilon} \\ &= \frac{p^2\lambda_2(1-\varepsilon)\,a_{\mathrm{L}}\,\left(1-\widetilde{L}^{-1}\left(1-\sqrt{1-\frac{P_{\mathrm{b}}}{p}}\right)\right)^2}{\varepsilon} \end{aligned} \tag{4.49}$$

where the last equality follows since $\frac{a_{\mathrm{R}}}{a_{\mathrm{L}}} = \frac{R}{1-R}$ (see Figure 4.1) and $R = (1-\varepsilon)C$. To continue the proof, we derive a lower bound on $1-\widetilde{L}^{-1}(x)$. Following the same steps which lead to (4.26) gives the inequality

$$1-\widetilde{L}^{-1}(x) \geq 1-\sqrt{\frac{x}{\widetilde{L}_2}}\,, \quad \forall\, x \geq 0 \tag{4.50}$$

where (4.32) implies that

$$\widetilde{L}_2 = \frac{\widetilde{L}''(0)}{2} = \frac{p\,L''(0)}{2} = p\,L_2\,. \tag{4.51}$$

Under the assumption that $1-\sqrt{1-\frac{P_{\mathrm{b}}}{p}} < p\,L_2$, substituting (4.50) and (4.51) in (4.49) gives

$$l \geq \frac{p\lambda_2(1-\varepsilon)\,a_{\mathrm{L}}\,\left(\sqrt{p\,L_2}-\sqrt{1-\sqrt{1-\frac{P_{\mathrm{b}}}{p}}}\right)^2}{L_2\,\varepsilon}\,. \tag{4.52}$$

Finally, the lower bound on the number of iterations in (4.8) follows from (4.52) by substituting $L_2 = \frac{\lambda_2\,a_{\mathrm{L}}}{2}$.

Considering the case where $P_{\mathrm{b}} \to 0$ on both sides of (4.8) gives

$$l(\varepsilon, p, P_{\mathrm{b}} \to 0) \geq 2p^2\,(1-\varepsilon)\,\frac{L_2(\varepsilon)}{\varepsilon}\,.$$

4.5 Summary and Conclusions

In this paper, we consider the number of iterations which is required for successful message-passing decoding of code ensembles defined on graphs. In the considered setting, we let the block length of these ensembles tend to infinity, and the transmission takes place over a binary erasure channel (BEC).

In order to study the decoding complexity of these code ensembles under iterative decoding, one needs also to take into account the graphical complexity of the Tanner graphs of these code ensembles. For the BEC, this graphical complexity is closely related to the total number of operations performed by the iterative decoder. For various families of code ensembles, Table 4.1 compares the number of iterations and the graphical complexity which are required to achieve a given fraction $1 - \varepsilon$ (where ε can be made arbitrarily small) of the capacity of a BEC with vanishing bit erasure probability. The results in Table 4.1 are based on lower bounds and some achievability results which are related to the graphical complexity of various families of code ensembles defined on graphs (see [65, 64, 81, 80]); the results related to the number of iterations are based on the lower bounds derived in this paper.

Code family	Number of iterations as function of ε	Graphical complexity as function of ε
LDPC	$\Omega\left(\frac{1}{\varepsilon}\right)$ (Theorem 4.1)	$\Theta\left(\ln \frac{1}{\varepsilon}\right)$ [81, Theorems 2.1 and 2.3]
Systematic IRA	$\Omega\left(\frac{1}{\varepsilon}\right)$ (Theorem 4.3)	$\Theta\left(\ln \frac{1}{\varepsilon}\right)$ [80, Theorems 1 and 2]
Non-systematic IRA	$\Omega\left(\frac{1}{\varepsilon}\right)$ (Theorem 4.3)	$\Theta(1)$ [65]
Systematic ARA	$\Omega\left(\frac{1}{\varepsilon}\right)$ (Theorem 4.2)	$\Theta(1)$ [64]

Table 4.1: Number of iterations and graphical complexity required to achieve a fraction $1 - \varepsilon$ of the capacity of a BEC with vanishing bit erasure probability under iterative message-passing decoding.

Theorems 4.1–4.3 demonstrate that for various attractive families of code ensembles (including low-density parity-check (LDPC) codes, systematic and non-systematic irregular repeat-accumulate (IRA) codes, and accumulate-repeat-accumulate (ARA) codes), the number of iterations which is required to achieve a desired bit erasure probability scales at least like the inverse of the gap between the channel capacity and the design rate of the ensemble. This conclusion holds provided that the fraction of degree-2 variable nodes in the Tanner graph does not tend to zero as the gap to capacity vanishes (where under mild conditions, this property is satisfied for sequences of capacity-achieving LDPC ensembles, see [76, Lemma 5]).

When the graphical complexity of these families of ensembles is considered, the results are less homogenous. More explicitly, assume a sequence of LDPC codes (or ensembles) whose block length tends to infinity, and consider the case where their transmission takes place over a memoryless binary-input output-symmetric channel. It follows from [81, Theorem 2.1] that if a fraction $1-\varepsilon$ of the capacity of this channel is achieved with vanishing bit error (erasure) probability under ML decoding (or any sub-optimal decoding algorithm), then the graphical complexity of an arbitrary representation of the codes using bipartite graphs scales at least like $\ln \frac{1}{\varepsilon}$. For systematic IRA codes which are transmitted over the BEC and decoded by a standard iterative message-passing decoder, a similar result on their graphical complexity is obtained in [80, Theorem 1]. In [81, Theorem 2.3], the lower bound on the graphical complexity of LDPC ensembles is achieved for the BEC (up to a small additive constant), even under iterative message-passing decoding, by the right-regular LDPC ensembles of Shokrollahi [94]. Similarly, [80, Theorem 2] presents an achievability result of this form for ensembles of systematic IRA codes transmitted over the BEC; the graphical complexity of these ensembles scales logarithmically with $\frac{1}{\varepsilon}$. For ensembles of non-systematic IRA and systematic ARA codes, however, the addition of state nodes in their standard representation by Tanner graphs allows to achieve an improved trade-off between the gap to capacity and the graphical complexity; suitable constructions of such ensembles enable to approach the capacity of the BEC with vanishing bit erasure probability under iterative decoding while maintaining *a bounded graphical complexity* (see [65] and [64]). We note that the ensembles in [64] have the additional advantage of being systematic, which allows simple decoding of the information bits.

The lower bounds on the number of iterations in Theorems 4.1–4.3 become trivial when the fraction of degree-2 variable nodes vanishes. As noted in Discussion 4.2, for all known capacity-approaching sequences of LDPC ensembles, this fraction tends to $\frac{1}{2}$ as the gap to capacity vanishes. For some ensembles of capacity approaching systematic ARA codes presented in [64], the fraction of degree-2 'punctured bit' nodes (as introduced in Figure 4.1) is defined to be zero (see [64, Table I]). For these ensembles, the lower bound on the number of iterations in Theorem 4.2 is ineffective. However, this is mainly a result of our focus on the derivation of simple lower bounds on the number of iterations which do not depend on the full characterization of the degree distributions of the code ensembles. Following the proofs of Theorems 4.1 and 4.2, and focusing on the case where the fraction of degree-2 variable nodes vanishes, it is possible to derive lower bounds on the number of iterations which are not trivial even in this case; these bounds, however, require the knowledge of the entire degree distribution of the examined ensembles.

The simple lower bounds on the number of iterations of graph-based ensembles, as derived in this paper, scale like the inverse of the gap in rate to capacity and also depend on the target bit erasure probability. The behavior of these lower bounds matches well with the experimental results and the conjectures on the number of iterations and complexity, as provided by Khandekar and McEliece (see [43, 42, 56]). In [42, Theorem 3.5], it was stated that for LDPC and IRA ensembles which achieve a fraction $1-\varepsilon$ of the channel capacity of a BEC with a target bit erasure probability of P_{b} under iterative message-passing decoding, the number of iterations grows like $O\left(\frac{1}{\varepsilon}\right)$. In light of the outline of the proof of this statement, as suggested in [42, p. 71], it implicitly assumes that the flatness condition is satisfied for these code ensembles and also that the target bit erasure probability vanishes; under these assumptions, the reasoning suggested by Khandekar in [42, Section 3.6] supports the behavior of the lower bounds which are derived in this paper.

The matching condition for generalized extrinsic information transfer (GEXIT) curves serves to conjecture in [59, Section XI] that the number of iterations scales like the inverse of the achievable gap in rate to capacity (see also [57, p. 92]); this conjecture refers to LDPC ensembles whose transmission takes place over a general memoryless binary-input output-symmetric (MBIOS) channel. Focusing on the BEC, the derivation of the lower bounds on the number of iterations (see Section 4.4) makes the heuristic reasoning of this scaling rigorous. It also extends the bounds to various graph-based code ensembles (e.g., IRA and ARA ensembles) under iterative message-passing decoding, and makes them universal for the BEC in the sense that they are expressed in terms of some basic parameters of the ensembles which include the fraction of degree-2 variable nodes, the target bit erasure probability and the asymptotic gap between the channel capacity and the design rate of the ensemble (but the bounds here do not depend explicitly on the degree distributions of the code ensembles). An interesting and challenging direction which calls for further research is to extend these lower bounds on the number of iterations for general MBIOS channels; as suggested in [59, Section XI], a consequence of the matching condition for GEXIT curves has the potential to lead to such lower bounds on the number of iterations which also scale like the inverse of the gap to capacity for general MBIOS channels.

Appendices

4.A Proof of Proposition 4.1

We begin the proof by considering an iterative decoder of systematic ARA codes by viewing them as interleaved and serially concatenated codes. The outer code of the

systematic ARA code consists of the first accumulator which operates on the systematic bits (see the upper zigzag in Figure 4.1), followed by the irregular repetition code. The inner code consists of the irregular SPC code, followed by the second accumulator (see the lower zigzag in Figure 4.1). These two constituent codes are joined by an interleaver which permutes the repeated bits at the output of the outer code before they are used as input to the inner encoder; for the considered ARA ensemble, we assume that the interleaver is chosen uniformly at random over all interleavers of the appropriate length. The turbo-like decoding algorithm is based on iterating extrinsic information between bitwise MAP decoders of the two constituent codes (see e.g., [11]). Each decoding iteration begins with an extrinsic bitwise MAP decoding for each non-systematic output bit of the outer code (these are the bits which serve as input to the inner code) based on the information regarding these bits received from the extrinsic bitwise MAP decoder of the inner code in the previous iteration and the information on the systematic bits received from the communication channel. In the second stage of the iteration, this information is passed from the outer decoder to an extrinsic bitwise MAP decoder of the inner code and is used as a-priori knowledge for decoding the input bits of the inner code. A Tanner graph for turbo-like decoding of systematic ARA codes is presented in Figure 4.3. Considering the asymptotic case where the block length tends to infinity, we denote the probability of erasure messages from the outer decoder to the inner decoder and vice versa at the l'th decoding iteration by $x_0^{(l)}$ and $x_1^{(l)}$, respectively. Keeping in line with the notation in the proofs of Theorems 4.1 and 4.2, we begin counting the iterations at $l = 0$. Since there is no a-priori information regarding the non-systematic output bits of the outer decoder (which are permuted to form the input bits of the inner decoder, as shown in Figure 4.3) we have

$$x_0^{(-1)} = x_1^{(-1)} = 1. \tag{4.A.1}$$

We now turn to calculate the erasure probability $x_0^{(l)}$ in an extrinsic bitwise MAP decoding of non-systematic output bits of the outer code, given that the a-priori erasure probability of these bits is $x_1^{(l-1)}$. To this end, we consider the Tanner graph of the outer code, shown in the top box of Figure 4.3. We note that this Tanner graph contains no cycles, and therefore bitwise MAP decoding of this code can be performed by using the standard iterative message-passing decoding algorithm until a fixed-point is reached. In such a decoder which operates on the Tanner graph of the outer code, messages are transferred between the 'punctured bit' and the 'parity-check 1' nodes of the graph. Let us denote by $x_{0,\mathrm{o}}(x)$ the probability of erasure in messages from the 'punctured bit' nodes to the 'parity-check 1' nodes at the fixed point of the iterative decoding algorithm, when the a-priori erasure probability of the

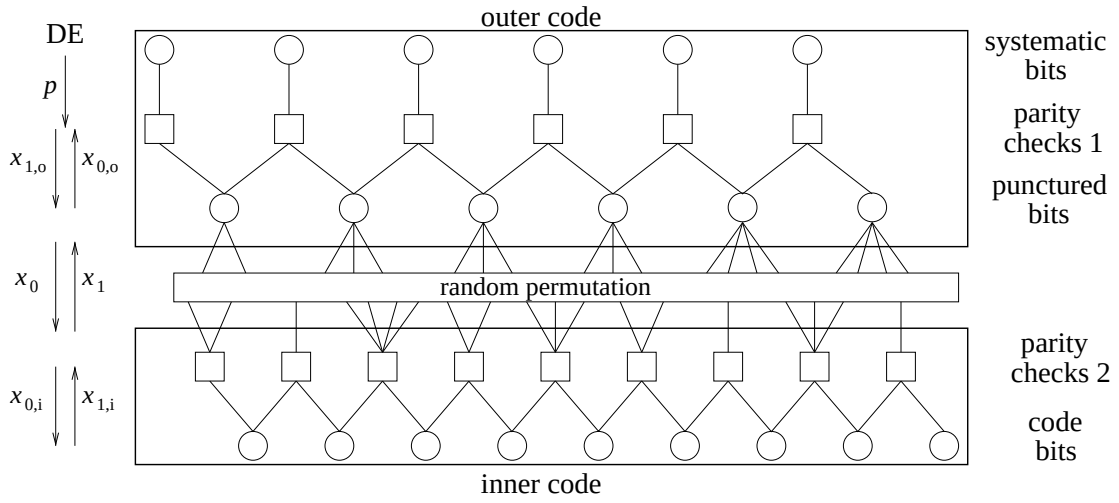

Figure 4.3: Tanner graph of a systematic accumulate-repeat-accumulate (ARA) code for turbo-like decoding as an interleaved and serially concatenated code.

output bits is x. Similarly, we denote by $x_{1,\mathrm{o}}(x)$ the erasure probability in messages from the 'parity-check 1' nodes to the 'punctured bit' nodes at the fixed point, where x is the a-priori erasure probability of the non-systematic output bits. Based on the structure of the Tanner graph, we have

$$x_{0,\mathrm{o}}(x) = x_{1,\mathrm{o}}(x) \cdot L(x) \tag{4.A.2}$$

and

$$x_{1,\mathrm{o}}(x) = 1 - (1-p)\big(1 - x_{0,\mathrm{o}}(x)\big) \tag{4.A.3}$$

where L is defined in (4.1) and it forms the degree distribution of the 'punctured bit' nodes from the node perspective, and p denotes the erasure probability of the BEC. Substituting (4.A.2) into (4.A.3) gives

$$x_{1,\mathrm{o}}(x) = \frac{p}{1-(1-p)L(x)}\,. \tag{4.A.4}$$

Therefore, the structure of the Tanner graph of the outer code implies that the erasure probability $x_0^{(l)}$ in messages from the outer decoder to the inner decoder at iteration number l of the turbo-like decoding algorithm is given by

$$\begin{aligned} x_0^{(l)} &= \Big(x_{1,\mathrm{o}}\big(x_1^{(l-1)}\big)\Big)^2 \lambda\big(x_1^{(l-1)}\big) \\ &= \left(\frac{p}{1-(1-p)L\big(x_1^{(l-1)}\big)}\right)^2 \lambda\big(x_1^{(l-1)}\big) \\ &= \widetilde{\lambda}\big(x_1^{(l-1)}\big) \end{aligned} \tag{4.A.5}$$

where the second equality relies on (4.A.4), and $\widetilde{\lambda}$ is the tilted degree distribution which results from graph reduction (see (4.29)). We now employ a similar technique to calculate the erasure probability $x_1^{(l)}$ in an extrinsic bitwise MAP decoding of input bits of the inner code, given that the a-priori erasure probability of these bits is $x_0^{(l)}$. Since the Tanner of the inner code is also cycle-free (see the lower box in Figure 4.3), extrinsic bitwise MAP decoding can be done by using the iterative decoder operating on the Tanner graph of the inner code. We denote by $x_{0,\mathrm{i}}(x)$ the erasure probability of messages from the 'parity check 2' nodes to the 'code bit' nodes at the fixed point of the iterative decoding algorithm when x is the a-priori erasure probability of the input bits. Similarly, $x_{1,\mathrm{i}}(x)$ designates the erasure probability of messages from the 'code bit' nodes to the 'parity check 2' nodes at the fixed point of the decoding algorithm, when x is the a-priori erasure probability of the input bits. The structure of the Tanner graph implies that

$$x_{0,\mathrm{i}}(x) = 1 - \big(1 - x_{1,\mathrm{i}}(x)\big) R(1-x) \tag{4.A.6}$$

and

$$x_{1,\mathrm{i}}(x) = p\, x_{0,\mathrm{i}}(x) \tag{4.A.7}$$

where R is defined in (4.2). Substituting (4.A.6) into (4.A.7) gives

$$x_{1,\mathrm{i}}(x) = \frac{p\big(1 - R(1-x)\big)}{1 - p\, R(1-x)}\,. \tag{4.A.8}$$

Therefore, the erasure probability $x_1^{(l)}$ in messages from the inner decoder to the outer decoder at iteration number l of the turbo-like decoding algorithm is given by

$$\begin{aligned} x_1^{(l)} &= 1 - \Big(1 - x_{1,\mathrm{i}}\big(x_0^{(l)}\big)\Big)^2 \rho\big(1 - x_0^{(l)}\big) \\ &= 1 - \left(1 - \frac{p\Big(1 - R\big(1 - x_0^{(l)}\big)\Big)}{1 - p\, R\big(1 - x_0^{(l)}\big)}\right)^2 \rho\big(1 - x_0^{(l)}\big) \\ &= 1 - \left(\frac{1-p}{1 - p\, R\big(1 - x_0^{(l)}\big)}\right)^2 \rho\big(1 - x_0^{(l)}\big) \\ &= 1 - \widetilde{\rho}\big(1 - x_0^{(l)}\big) \end{aligned} \tag{4.A.9}$$

where the second equality relies on (4.A.8), and $\widetilde{\rho}$ is the tilted degree distribution resulting from graph reduction (see (4.30)). Combining (4.A.1), (4.A.5) and (4.A.9) gives

$$\begin{aligned} x_0^{(0)} &= \widetilde{\lambda}\big(x_1^{(-1)}\big) = \widetilde{\lambda}(1) = 1\,, \\ x_0^{(l)} &= \widetilde{\lambda}\Big(1 - \widetilde{\rho}\big(1 - x_0^{(l-1)}\big)\Big)\,, \qquad l \in \mathbb{N}\,. \end{aligned} \tag{4.A.10}$$

Observing the proof of Theorem 4.2, we note that $x_0^{(l)} = x^{(l)}$ for all $l = 0, 1, \ldots$, where is the $x^{(l)}$ value at the left tip of the horizontal line h_l in Figure 4.2 (see Eq. (4.46) on page 148).

Let $P_{\mathrm{b}}^{(l)}$ designate the average erasure probability of the systematic bits at the end of the l'th iteration of the turbo-like decoder. From the definition of the turbo-like decoding algorithm, $P_{\mathrm{b}}^{(l)}$ is the erasure probability of bitwise MAP decoding for the input bits to the outer code, given that the a-priori erasure probability of the output bits of this code is given by $x_1^{(l)}$. Based of the structure of the Tanner graph of the outer code in Figure 4.3, we get

$$P_{\mathrm{b}}^{(l)} = p\,\left[1 - \Big(1 - x_{0,\mathrm{o}}\big(x_1^{(l)}\big)\Big)^2\right] \tag{4.A.11}$$

where $x_{0,\mathrm{o}}(x)$ in the fixed point erasure probability of messages from the 'punctured bit' nodes to the 'parity-check 1' nodes in the case that the a-priori erasure probability of the non-systematic output bits of the code is x. Substituting (4.A.3) in (4.A.2) gives

$$x_{0,\mathrm{o}}(x) = \frac{p\,L(x)}{1 - (1-p)L(x)}\,.$$

Substituting the above equality into (4.A.11), we have

$$\begin{aligned} P_{\mathrm{b}}^{(l)} &= p\,\left[1 - \left(1 - \frac{p\,L\big(x_1^{(l)}\big)}{1 - (1-p)L\big(x_1^{(l)}\big)}\right)^2\right] \\ &= p\,\left[1 - \Big(1 - \widetilde{L}\big(x_1^{(l)}\big)\Big)^2\right] \\ &= p\,\left[1 - \Big(1 - \widetilde{L}\Big(1 - \widetilde{\rho}\big(1 - x_0^{(l)}\big)\Big)\Big)^2\right] \end{aligned}$$

where the second equality follows from the definition of $\widetilde{L}$ in (4.32) and the third equality relies on (4.A.9). Using simple algebra, the above expression gives

$$1 - \sqrt{1 - \frac{P_{\mathrm{b}}^{(l)}}{p}} = \widetilde{L}\Big(1 - \widetilde{\rho}\big(1 - x_0^{(l)}\big)\Big)\,. \tag{4.A.12}$$

Hence, the lower bound on the average erasure probability of the systematic bits at the end of the l'th iteration of the standard iterative decoder for ARA codes in Lemma 4.3 is satisfied (with equality) also for the turbo-like decoder.

Let l designate the required number of iterations for the turbo-like decoder to achieve an average erasure probability P_{b} of the systematic bits. Since we start counting the iterations at zero, (4.A.12) implies that l is the smallest natural number which satisfies

$$1-\sqrt{1-\frac{P_{\mathrm{b}}}{p}} \geq \widetilde{L}\Big(1-\widetilde{\rho}\big(1-x_0^{(l-1)}\big)\Big)\,.$$

However, this is exactly the quantity for which we calculated the lower bound in the proof of Theorem 4.2 (see Lemmas 4.2 and 4.3 and Eq. (4.31)). Therefore, we conclude that the lower bound on the number of iterations (l) in Theorem 4.2 holds also when the considered turbo-like decoding algorithm is employed to decode the systematic ARA codes as interleaved and serially concatenated codes.

4.B Some mathematical details related to the proof of Theorem 4.2

4.B.1 Proof of Lemma 4.2

The proof of Lemma 4.2 is based on the DE equations in (4.5) for systematic ARA ensembles. From the DE equations for $x_2^{(l)}$ and $x_3^{(l)}$, we have

$$\begin{aligned} x_3^{(l)} &= p\, x_2^{(l)} \\ &= p\left[1-R\left(1-x_1^{(l)}\right)\left(1-x_3^{(l-1)}\right)\right] \\ &\geq p\left[1-R\left(1-x_1^{(l)}\right)\left(1-x_3^{(l)}\right)\right] \end{aligned}$$

where the inequality follows since the decoding process does not add erasures, so $x_i^{(l)}$ is monotonically decreasing with l (for $i=0,1,\ldots,5$). This gives

$$1-x_3^{(l)} \leq 1-p\left[1-R\left(1-x_1^{(l)}\right)\left(1-x_3^{(l)}\right)\right]$$

and

$$1-x_3^{(l)} \leq \frac{1-p}{1-pR\left(1-x_1^{(l)}\right)}. \tag{4.B.1}$$

Substituting (4.B.1) into the DE equation for $x_4^{(l)}$ (see (4.5)) gives

$$\begin{aligned} x_4^{(l)} &= 1-\left(1-x_3^{(l)}\right)^2 \rho\left(1-x_1^{(l)}\right) \\ &\geq 1-\left(\frac{1-p}{1-pR\left(1-x_1^{(l)}\right)}\right)^2 \rho\left(1-x_1^{(l)}\right) \\ &= 1-\widetilde{\rho}\left(1-x_1^{(l)}\right) \end{aligned} \tag{4.B.2}$$

where $\widetilde{\rho}$ is defined in (4.30). From (4.5), we get

$$\begin{aligned} x_5^{(l)} &= x_0^{(l)} L\left(x_4^{(l)}\right) \\ &= \left[1-\left(1-x_5^{(l-1)}\right)(1-p)\right] L\left(x_4^{(l)}\right) \\ &\geq \left[1-\left(1-x_5^{(l)}\right)(1-p)\right] L\left(x_4^{(l)}\right) \end{aligned}$$

where the inequality follows from the monotonicity of $\{x_5^{(l)}\}$. Solving for $1-x_5^{(l)}$ gives

$$1-x_5^{(l)} \leq \frac{1-L\left(x_4^{(l)}\right)}{1-(1-p)L\left(x_4^{(l)}\right)}. \tag{4.B.3}$$

Substituting (4.B.3) into the DE equation for $x_0^{(l)}$ in (4.5), we have

$$\begin{aligned} x_0^{(l)} &= 1-\left(1-x_5^{(l-1)}\right)(1-p) \\ &\geq 1-\frac{(1-p)\left[1-L\left(x_4^{(l-1)}\right)\right]}{1-(1-p)L\left(x_4^{(l-1)}\right)} \\ &= \frac{p}{1-(1-p)L\left(x_4^{(l-1)}\right)}. \end{aligned}$$

Substituting the inequality above into the DE equation for $x_1^{(l)}$ gives

$$\begin{aligned} x_1^{(l)} &= \left(x_0^{(l)}\right)^2 \lambda\left(x_4^{(l-1)}\right) \\ &\geq \left(\frac{p}{1-(1-p)L\left(x_4^{(l-1)}\right)}\right)^2 \lambda\left(x_4^{(l-1)}\right) \\ &= \widetilde{\lambda}\left(x_4^{(l-1)}\right) \end{aligned} \tag{4.B.4}$$

where $\widetilde{\lambda}$ is defined in (4.29). Finally (4.28) follows from (4.B.2) and (4.B.4) and the monotonicity of $\widetilde{\lambda}$ over the interval $[0,1]$.

4.B.2 Proof of Lemma 4.3

From the structure of the Tanner graph of systematic ARA codes (see Figure 4.1) and the DE equation for $x_5^{(l)}$ in (4.5) we get

$$\begin{aligned} P_{\mathrm{b}}^{(l)} &= p\left[1-\left(1-x_5^{(l)}\right)^2\right] \\ &= p\left[1-\left(1-x_0^{(l)}L\left(x_4^{(l)}\right)\right)^2\right]. \end{aligned} \tag{4.B.5}$$

The DE equation (4.5) for $x_1^{(l)}$ and (4.29) imply that

$$\begin{aligned} \left(x_0^{(l)}\right)^2 &= \frac{x_1^{(l)}}{\lambda\left(x_4^{(l-1)}\right)} \\ &= \frac{x_1^{(l)}\,p^2}{\widetilde{\lambda}\left(x_4^{(l-1)}\right)\,\left[1-(1-p)\,L\left(x_4^{(l-1)}\right)\right]^2} \\ &\geq \left(\frac{p}{1-(1-p)\,L\left(x_4^{(l-1)}\right)}\right)^2 \end{aligned}$$

where the last inequality follows from (4.B.4). Taking the square root on both sides of the above inequality gives

$$x_0^{(l)} \geq \frac{p}{1-(1-p)\,L\left(x_4^{(l-1)}\right)}. \tag{4.B.6}$$

Substituting (4.B.6) in (4.B.5), we get

$$\begin{aligned} P_{\mathrm{b}}^{(l)} &\geq p\left[1-\left(1-\frac{p\,L\left(x_4^{(l)}\right)}{1-(1-p)\,L\left(x_4^{(l-1)}\right)}\right)^2\right] \\ &\geq p\left[1-\left(1-\frac{p\,L\left(x_4^{(l)}\right)}{1-(1-p)\,L\left(x_4^{(l)}\right)}\right)^2\right] \end{aligned} \tag{4.B.7}$$

where the second inequality above follows since the decoding process does not add erasures so $x_4^{(l)} \leq x_4^{(l-1)}$, and from the monotonicity of L over $[0,1]$. Applying the definition of $\widetilde{L}$ in (4.32) to the RHS of (4.B.7) gives

$$\begin{aligned} P_{\mathrm{b}}^{(l)} &\geq p\left[1-\left(1-\widetilde{L}\left(x_4^{(l)}\right)\right)^2\right] \\ &\geq p\left\{1-\left[1-\widetilde{L}\left(1-\rho\left(x_1^{(l)}\right)\right)\right]^2\right\} \end{aligned} \tag{4.B.8}$$

where the last inequality follows from (4.B.2). Finally, (4.34) follows directly from (4.B.8).

Chapter 5

An Improved Sphere-Packing Bound for Finite-Length Codes over Symmetric Memoryless Channels

This chapter is a preprint of

- G. Wiechman and I. Sason, "An improved sphere-packing bound for finite-length codes on symmetric memoryless channels," submitted to *IEEE Trans. on Information Theory*, March 2007.

Chapter Overview: This paper derives an improved sphere-packing (ISP) bound for finite-length error-correcting codes whose transmission takes place over symmetric memoryless channels, and the codes are decoded with an arbitrary list decoder. We first review classical results, i.e., the 1959 sphere-packing (SP59) bound of Shannon for the Gaussian channel, and the 1967 sphere-packing (SP67) bound of Shannon et al. for discrete memoryless channels. An improvement on the SP67 bound, as suggested by Valembois and Fossorier, is also discussed. These concepts are used for the derivation of a new lower bound on the error probability of list decoding (referred to as the ISP bound) which is uniformly tighter than the SP67 bound and its improved version. The ISP bound is applicable to symmetric memoryless channels, and some of its applications are exemplified. Its tightness under ML decoding is studied by comparing the ISP bound to previously reported upper and lower bounds on the ML decoding error probability, and also to computer simulations of iteratively decoded turbo-like codes. This paper also presents a technique which performs the entire calculation of the SP59 bound in the logarithmic domain, thus facilitating the exact calculation of this bound for moderate to large block lengths without the need for the asymptotic approximations provided by Shannon.

5.1 Introduction

The theoretical study of the fundamental performance limitations of long block codes was initiated by Shannon. During the fifties and sixties, this research work attracted Shannon and his colleagues at MIT and Bell Labs (see, e.g., the collected papers of Shannon [96] and the book of Gallager [32]). An overview of these classical results and their impact was addressed by Berlekamp [13].

The 1959 sphere-packing (SP59) bound of Shannon [89] serves for the evaluation of the performance limits of block codes whose transmission takes place over an AWGN channel. This lower bound on the decoding error probability is expressed in terms of the block length and rate of the code; however, it does not take into account the modulation used, but only assumes that the signals are of equal energy. It is often used as a reference for quantifying the sub-optimality of error-correcting codes under some practical decoding algorithms.

The 1967 sphere-packing (SP67) bound, derived by Shannon, Gallager and Berlekamp [87], provides a lower bound on the decoding error probability of block codes as a function of their block length and code rate, and applies to arbitrary discrete memoryless channels. Like the random-coding bound (RCB) of Gallager [31], the SP67 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 known to be tight at the portion of the rate region between the critical rate (R_{c}) and the channel capacity; for all the rates in this range, the error exponents of the SP67 bound and the RCB coincide (see [87, Part 1]).

The introduction of turbo-like codes, which closely approach the Shannon capacity limit with moderate block lengths and a feasible decoding complexity, stirred up new interest in studying the limits of code performance as a function of the block length (see, e.g., [27, 44, 45, 55, 79, 98, 109, 115]). In a recent paper [20], Costello and Forney survey the evolution of channel coding techniques, and also address the significant contributions of error-correcting codes in improving the tradeoff between performance, block length (delay) and complexity for practical applications.

In spite of the exponential decay of the SP67 bound in terms of the block length at all rates below the channel capacity, this bound appears to be loose for codes of small to moderate block lengths. The weakness of this bound is due to the original focus in [87] on asymptotic analysis. In [109], Valembois and Fossorier revisited the SP67 bound in order to improve its tightness for finite-length block codes (especially, for codes of short to moderate block lengths), and also extended its validity to memoryless continuous-output channels (e.g., the binary-input AWGN channel). The remarkable

improvement of their bound over the classical SP67 bound was exemplified in [109]. Moreover, the extension of the bound in [109] to memoryless continuous-output channels provides an alternative to the SP59 bound which holds for the AWGN channel [89].

This paper is focused on the study of the fundamental performance limitations of finite-length error-correcting codes and the tradeoff between their performance and block length when the transmission takes place over an arbitrary symmetric memoryless channel. This study is facilitated by theoretical bounds, and it is also compared to the performance of modern coding techniques under sub-optimal and practical decoding algorithms. In this work, we derive an improved sphere-packing bound (referred to as the 'ISP bound') which improves the bounding techniques in [87] and [109], especially for codes of short to moderate block lengths; this new bound is valid for all symmetric memoryless channels.

The structure of this paper is as follows: Section 5.2 reviews the concepts used in the derivation of the SP67 bound [87, Part 1] and its improved version in [109]. In Section 5.3, we derive the ISP bound which improves the bound in [109] for symmetric memoryless channels where the derivation of the ISP bound relies on concepts and notation presented in Section 5.2. Section 5.4 starts by reviewing the SP59 bound of Shannon [89], and presenting an algorithm used in [109] for a numerical calculation this bound. The numerical instability of this algorithm for codes of moderate to large block lengths motivates the derivation of an alternative algorithm in Section 5.4 which facilitates the exact calculation of the SP59 bound, irrespectively of the block length. Section 5.5 provides numerical results which serve to compare the ISP bound to previously reported sphere-packing bounds. The tightness of the ISP bound is exemplified in Section 5.5 for various communication channels. Additionally, sphere-packing bounds are applied in Section 5.5 to study the tradeoff between the performance and the required block length of error-correcting codes. We conclude our discussion in Section 5.6. Some technical details are relegated to the appendices.

5.2 The 1967 Sphere-Packing Bound and Improvements

In the following, we present the SP67 bound and its improvement in [109], followed by an outline of their derivation. Classical sphere-packing bounds are reviewed in [79, Chapter 5]. This section serves as a preparatory step towards the derivation of an improved sphere-packing bound in the next section.

5.2.1 The 1967 Sphere-Packing Bound

Let us consider a block code $\mathcal{C}$ which consists of M codewords each of length N, and denote its codewords by $\mathbf{x_1}, \ldots, \mathbf{x_M}$. Assume that $\mathcal{C}$ is transmitted over a discrete memoryless channel (DMC) and is decoded by a *list decoder*; for each received sequence $\mathbf{y}$, the decoder outputs a list of at most L integers from the set $\{1, 2, \ldots, M\}$ which correspond to the indices of the codewords. A list decoding error is declared if the index of the transmitted codeword does not appear in the list. Originally introduced by Elias [29] and Wozencraft [120], list decoding signifies an important class of decoding algorithms. During the last decade, there has been a significant breakthrough in the construction of efficient list-decoding algorithms for error-correcting codes (see, e.g., [33], [75, Chapter 9] and references therein).

A lower bound on the decoding error probability of an arbitrary block code with M codewords of length N is derived in [87]. This bound applies to an arbitrary list decoder where the size of the list is limited to L. The particular case where $L = 1$ clearly provides a lower bound on the error probability under maximum-likelihood (ML) decoding.

Let $\mathcal{Y}_m$ designate the set of output sequences $\mathbf{y}$ for which message m is on the decoding list, and define $P_m(\mathbf{y}) \triangleq \Pr(\mathbf{y}|\mathbf{x_m})$. The conditional error probability under list decoding when message m is sent over the channel is given by

$$P_{\mathrm{e},m} = \sum_{y \in \mathcal{Y}_m^{\mathrm{c}}} P_m(\mathbf{y}) \tag{5.1}$$

where the superscript 'c' stands for the complementary set. For the block code and list decoder under consideration, let $P_{\mathrm{e,max}}$ designate the maximal value of $P_{\mathrm{e},m}$ where $m \in \{1, 2, \ldots, M\}$. Assuming that all the codewords are equally likely to be transmitted, the average decoding error probability is given by

$$P_{\mathrm{e}} = \frac{1}{M} \sum_{m=1}^{M} P_{\mathrm{e},m}.$$

Referring to a list decoder of size at most L, the code rate is defined as $R \triangleq \frac{\ln\left(\frac{M}{L}\right)}{N}$ nats per channel use.

The derivation of the SP67 bound [87, Part 1] is divided into three main steps. The first step refers to the derivation of upper and lower bounds on the error probability of a code consisting of two codewords only. These bounds are given by the following theorem:

Theorem 5.1 [Upper and Lower Bounds on the Pairwise Error Probability] [87, Theorem 5]. Let P_1 and P_2 be two probability assignments defined over a discrete set of sequences $\mathcal{Y}$, $\mathcal{Y}_1$ and $\mathcal{Y}_2 = \mathcal{Y}_1^c$ be (disjoint) decision regions for these sequences, $P_{e,1}$ and $P_{e,2}$ be given by (5.1), and assume that $P_1(\mathbf{y})P_2(\mathbf{y}) \neq 0$ for at least one sequence $\mathbf{y}$. Then, for all $s \in (0,1)$

$$P_{e,1} > \frac{1}{4}\exp\Big(\mu(s) - s\mu'(s) - s\sqrt{2\mu''(s)}\Big) \tag{5.2}$$

or

$$P_{e,2} > \frac{1}{4}\exp\Big(\mu(s) + (1-s)\mu'(s) - (1-s)\sqrt{2\mu''(s)}\Big) \tag{5.3}$$

where

$$\mu(s) \triangleq \ln\Big(\sum_{\mathbf{y}} P_1(\mathbf{y})^{1-s} P_2(\mathbf{y})^s\Big), \qquad 0 < s < 1. \tag{5.4}$$

Furthermore, for an appropriate choice of the decision regions $\mathcal{Y}_1$ and $\mathcal{Y}_2$, the following upper bounds hold:

$$P_{e,1} \leq \exp\Big(\mu(s) - s\mu'(s)\Big) \tag{5.5}$$

and

$$P_{e,2} \leq \exp\Big(\mu(s) + (1-s)\mu'(s)\Big). \tag{5.6}$$

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 $\mathbf{y}$.

In the following, we present an outline of the proof of Theorem 5.1 which serves to emphasize the parallelism between Theorem 5.1 and the first part of the derivation of the ISP bound in Section 5.3. A detailed proof of this theorem is given in [87, Section III].

Proof: Let us define the log-likelihood ratio (LLR) as

$$D(\mathbf{y}) \triangleq \ln\left(\frac{P_2(\mathbf{y})}{P_1(\mathbf{y})}\right) \tag{5.7}$$

and the probability distribution

$$Q_s(\mathbf{y}) \triangleq \frac{P_1(\mathbf{y})^{1-s} P_2(\mathbf{y})^s}{\sum_{\mathbf{y}'} P_1(\mathbf{y}')^{1-s} P_2(\mathbf{y}')^s}, \quad 0 < s < 1. \tag{5.8}$$

It is simple to show that for all $0 < s < 1$, the first and second derivatives of μ in (5.4) are equal to the statistical expectation and variance of the LLR, respectively,

taken with respect to (w.r.t.) the probability distribution Q_s in (5.8). This gives the following equalities:

$$\mu'(s) = \mathbb{E}_{Q_s}\big(D(\mathbf{y})\big) \tag{5.9}$$

$$\mu''(s) = \mathrm{Var}_{Q_s}\big(D(\mathbf{y})\big). \tag{5.10}$$

Also, as can be readily verified from (5.4), (5.7) and (5.8)

$$P_1(\mathbf{y}) = \exp\big(\mu(s) - sD(\mathbf{y})\big)\, Q_s(\mathbf{y}) \tag{5.11}$$

$$P_2(\mathbf{y}) = \exp\big(\mu(s) + (1-s)D(\mathbf{y})\big)\, Q_s(\mathbf{y}). \tag{5.12}$$

For $0 < s < 1$, the equalities in (5.9) and (5.10) motivate the definition of a set of typical sequences w.r.t. the probability distribution Q_s as followed:

$$\mathcal{Y}_s \triangleq \left\{\mathbf{y} \in \mathcal{Y} : |D(\mathbf{y}) - \mu'(s)| \leq \sqrt{2\mu''(s)}\right\}. \tag{5.13}$$

For any choice of a decision region $\mathcal{Y}_1$, the conditional error probability given that the first message was transmitted satisfies

$$\begin{aligned} P_{\mathrm{e},1} &= \sum_{\mathbf{y}\in\mathcal{Y}_1^{\mathrm{c}}} P_1(\mathbf{y}) \\ &\geq \sum_{\mathbf{y}\in\mathcal{Y}_1^{\mathrm{c}}\bigcap\mathcal{Y}_s} P_1(\mathbf{y}) \\ &\overset{(a)}{=} \sum_{\mathbf{y}\in\mathcal{Y}_1^{\mathrm{c}}\bigcap\mathcal{Y}_s} \exp\Big(\mu(s) - sD(\mathbf{y})\Big)\, Q_s(\mathbf{y}) \\ &\overset{(b)}{\geq} \exp\Big(\mu(s) - s\mu'(s) - s\sqrt{2\mu''(s)}\Big) \sum_{\mathbf{y}\in\mathcal{Y}_1^{\mathrm{c}}\bigcap\mathcal{Y}_s} Q_s(\mathbf{y}) \end{aligned} \tag{5.14}$$

where (a) follows from (5.11) and (b) relies on the definition of $\mathcal{Y}_s$ in (5.13). Using similar arguments and relying on (5.12), we also get that

$$P_{\mathrm{e},2} \geq \exp\Big(\mu(s) + (1-s)\mu'(s) - (1-s)\sqrt{2\mu''(s)}\Big) \sum_{\mathbf{y}\in\mathcal{Y}_2^{\mathrm{c}}\bigcap\mathcal{Y}_s} Q_s(\mathbf{y})\,. \tag{5.15}$$

Since $\mathcal{Y}_1$ and $\mathcal{Y}_2$ form a partition of the observation space, we have that

$$\sum_{\mathbf{y}\in\mathcal{Y}_1^{\mathrm{c}}\bigcap\mathcal{Y}_s} Q_s(\mathbf{y}) + \sum_{\mathbf{y}\in\mathcal{Y}_2^{\mathrm{c}}\bigcap\mathcal{Y}_s} Q_s(\mathbf{y}) = \sum_{\mathbf{y}\in\mathcal{Y}_s} Q_s(\mathbf{y}) > \frac{1}{2}$$

where the last transition relies on (5.9), (5.10) and (5.13), and it follows from Chebychev's inequality. Therefore, at least one of the two sums on the LHS of the expression above must be greater than $\frac{1}{4}$. Substituting this in (5.14) and (5.15) completes the

proof on the satisfiability of at least one of the inequalities (5.2) and (5.3). The upper bound on the error probability in (5.5) and (5.6) is attained by selecting the decision region for the first codeword to be

$$\mathcal{Y}_1 \triangleq \big\{\mathbf{y} \in \mathcal{Y} : D(\mathbf{y}) < \mu'(s)\big\}$$

and the decision region for the second code as $\mathcal{Y}_2 \triangleq \mathcal{Y}_1^{\mathrm{c}}$. The proof for the upper bounds in (5.5) and (5.6) follows directly from (5.11), (5.12) and the particular choice of $\mathcal{Y}_1$ and $\mathcal{Y}_2$ as above. ■

The initial motivation of Theorem 5.1 is the calculation of lower bounds on the error probability of a two-word code. Note that this theorem is valid for any pair of probability assignments P_1 and P_2 and decision regions $\mathcal{Y}_1$ and $\mathcal{Y}_2$ which form a partition of the observation space.

In the continuation of the derivation of the SP67 bound in [87], this theorem is used in order to control the size of a decision region of a particular codeword without directly referring to the other codewords. To this end, an arbitrary probability tilting measure f_N is introduced in [87] over all N-length sequences of channel outputs, requiring that it is factorized in the form

$$f_N(\mathbf{y}) = \prod_{n=1}^{N} f(y_n) \tag{5.16}$$

for an arbitrary output sequence $\mathbf{y} = (y_1, \ldots, y_N)$. The size of the set $\mathcal{Y}_m$ is defined as

$$F(\mathcal{Y}_m) \triangleq \sum_{\mathbf{y} \in \mathcal{Y}_m} f_N(\mathbf{y}). \tag{5.17}$$

Next, [87] relies on Theorem 5.1 in order to relate the conditional error probability $P_{\mathrm{e},m}$ and $F(\mathcal{Y}_m)$ for fixed composition codes; this is done by associating $\Pr(\cdot|\mathbf{x_m})$ and f_N with P_1 and P_2, respectively. Theorem 5.1 is applied to derive a parametric lower bound on the size of the decision region $\mathcal{Y}_m$ or on the conditional error probability $P_{\mathrm{e},m}$. Due to the fact that the list size is limited to L, then

$$\sum_{m=1}^{M} F(\mathcal{Y}_m) = \sum_{m=1}^{M} \sum_{\mathbf{y} \in \mathcal{Y}_m} f_N(\mathbf{y}) \leq L$$

since for every sequence $\mathbf{y}$, the relation $\mathbf{y} \in \mathcal{Y}_m$ holds for at most L indices $m \in \{1, \ldots, M\}$, and $\sum_{\mathbf{y}} f_N(\mathbf{y}) = 1$. Therefore, there exists an index m so that $F(\mathcal{Y}_m) \leq \frac{L}{M}$ and for this unknown value of m, one can upper bound the conditional error probability $P_{\mathrm{e},m}$ by

$$P_{\mathrm{e,max}} \triangleq \max_{m \in \{1,\ldots,M\}} P_{\mathrm{e},m}.$$

Using Theorem 5.1 with the above setting for the probability assignments P_1 and P_2, then $P_{\mathrm{e},1}$ and $P_{\mathrm{e},2}$ on the LHS of (5.2) and (5.3) are respectively replaced by $P_{\mathrm{e},m}$ and $F(\mathcal{Y}_m)$. For the above unknown value of m, whose existence is assured to be in the set $\{1,\ldots,M\}$, one can replace $P_{\mathrm{e},m}$ and $F(\mathcal{Y}_m)$ on the LHS of (5.2) and (5.3) by their upper bounds $P_{\mathrm{e,max}}$ and $\frac{L}{M}$, respectively. This provides a lower bound on $P_{\mathrm{e,max}}$ as long as the inequality which follows from the replacement of $F(\mathcal{Y}_m)$ by its upper bound $\left(\frac{L}{M}\right)$ on the LHS of (5.3) does not hold. Next, the probability assignment $f \triangleq f_s$ is optimized in [87], so as to get the tightest (i.e., maximal) lower bound on $P_{\mathrm{e,max}}$ within this form while considering a code whose composition minimizes the bound (so that the bound holds for all fixed composition codes). A solution for this min-max problem, as provided in [87, Eqs. (4.18)–(4.20)], leads to the following theorem which gives a lower bound on the maximal decoding error probability of an arbitrary fixed composition block code (for a more detailed review of these concepts, see [79, Section 5.3]).

Theorem 5.2 [Sphere-Packing Bound on the Maximal Error Probability of Fixed Composition Codes] [87, Theorem 6]. Let $\mathcal{C}$ be a *fixed composition block code* of M codewords and length N. Assume that the transmission of $\mathcal{C}$ takes place over a DMC, and let $P(j|k)$ be the set of transition probabilities characterizing this channel (where $j \in \{0,\ldots,J-1\}$ and $k \in \{0,\ldots,K-1\}$ designate the channel output and input, respectively). For an arbitrary list decoder where the size of the list is limited to L, the *maximal error probability* ($P_{\mathrm{e,max}}$) satisfies

$$P_{\mathrm{e,max}} \geq \exp\left[-N\left(E_{\mathrm{sp}}\Big(R - \frac{\ln 4}{N} - \varepsilon\Big) + \sqrt{\frac{8}{N}}\ln\Big(\frac{e}{\sqrt{P_{\min}}}\Big) + \frac{\ln 4}{N}\right)\right]$$

where $R \triangleq \frac{\ln\left(\frac{M}{L}\right)}{N}$ is the rate of the code, $P_{\min}$ designates the smallest non-zero transition probability of the DMC, the parameter ε is an arbitrarily small positive number, and the function E_{sp} is given by

$$E_{\mathrm{sp}}(R) \triangleq \sup_{\rho\geq 0}\big(E_0(\rho) - \rho R\big) \tag{5.18}$$

$$E_0(\rho) \triangleq \max_{\mathbf{q}} E_0(\rho,\mathbf{q}) \tag{5.19}$$

$$E_0(\rho,\mathbf{q}) \triangleq -\ln\left(\sum_{j=0}^{J-1}\Big[\sum_{k=0}^{K-1} q_k P(j|k)^{\frac{1}{1+\rho}}\Big]^{1+\rho}\right). \tag{5.20}$$

The maximum in the RHS of (5.19) is taken over all probability vectors $\mathbf{q} = (q_0,\ldots,q_{K-1})$, i.e., over all $\mathbf{q}$ with K non-negative components summing to 1.

The reason for considering fixed composition codes in [87] is that, in general, the optimal probability distribution f_s may depend on the composition of the codewords through the choice of the parameter s in $(0, 1)$ (see [87, p. 96]).

The next step in the derivation of the SP67 bound is the application of Theorem 5.2 to obtain a lower bound on the maximal decoding error probability of an arbitrary block code. This is performed by lower bounding the maximal decoding error probability of a block code by the maximal error probability of its largest fixed composition subcode. Since the number of possible compositions is polynomial in the block length, one can lower bound the rate of the largest fixed composition subcode by $R - O\left(\frac{\ln N}{N}\right)$ where R is the rate of the original code. Clearly, the rate loss caused by considering this subcode vanishes when the block length tends to infinity; however, it loosens the bound for codes of short to moderate block lengths. Finally, the bound on the maximal block error probability is transformed into a bound on the average block error probability by considering an expurgated code which contains half of the codewords of the original code with the lowest conditional error probability. This finally leads to the SP67 bound in [87, Part 1].

Theorem 5.3 [The 1967 Sphere-Packing Bound for Discrete Memoryless Channels] [87, 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 $k \in \{0, \ldots, K-1\}$ and $j \in \{0, \ldots, J-1\}$ 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 consider an arbitrary list decoder where the size of the list is limited to L. Then, the *average decoding error probability* of the code $\mathcal{C}$ satisfies

$$P_{\mathrm{e}}(N, M, L) \geq \exp\left\{-N\left[E_{\mathrm{sp}}\left(R - O_1\left(\frac{\ln N}{N}\right)\right) + O_2\left(\frac{1}{\sqrt{N}}\right)\right]\right\}$$

where $R \triangleq \frac{\ln\left(\frac{M}{L}\right)}{N}$, and the error exponent $E_{\mathrm{sp}}(R)$ is introduced in (5.18). The terms

$$\begin{aligned} O_1\left(\frac{\ln N}{N}\right) &= \frac{\ln 8}{N} + \frac{K \ln N}{N} \\ O_2\left(\frac{1}{\sqrt{N}}\right) &= \sqrt{\frac{8}{N}} \ln\left(\frac{e}{\sqrt{P_{\min}}}\right) + \frac{\ln 8}{N} \end{aligned} \tag{5.21}$$

scale like $\frac{\ln N}{N}$ and $\frac{1}{\sqrt{N}}$, respectively (hence, they both vanish as we let N tend to infinity), and $P_{\min}$ denotes the smallest non-zero transition probability of the DMC.

5.2.2 Recent Improvements on the 1967 Sphere-Packing Bound

In [109], Valembois and Fossorier revisited the derivation of the SP67 bound, focusing on finite-length block codes. They presented four modifications to the classical derivation in [87] which improve the pre-exponent of the SP67 bound. The new bound derived in [109] is also valid for memoryless channels with discrete input and continuous output (as opposed to the SP67 bound which is only valid for DMCs). In this section, we outline the improvements suggested in [109] and present the resulting bound.

The first modification suggested in [109] is the addition of a free parameter in the derivation of the lower bound on the decoding error probability of two-word codes; this free parameter is used in conjunction with Chebychev's inequality, and it is optimized in order to tighten the lower bounds on $P_{\mathrm{e},1}$ and $P_{\mathrm{e},2}$ in Theorem 5.1 (see (5.2), (5.3)).

A second improvement presented in [109] is related to the inequality $s\sqrt{\mu''(s)} \leq \ln\left(\frac{e}{\sqrt{P_{\min}}}\right)$ which was applied to simplify the final form of the bound in Theorem 5.3 (see [87, Part 1]). This bound on the second derivative of μ results in no asymptotic loss, but it loosens the lower bound on the decoding error probability for finite-length codes (especially, for short to moderate block lengths). By using the exact value of μ'' instead, the tightness of the resulting bound is further improved in [109]. This modification also makes the bound suitable to memoryless channels with a continuous output alphabet, as it is no longer required that $P_{\min}$ is positive. It should be noted that this causes a small discrepancy in the derivation of the bound; the derivation of a lower bound on the block error probability which is *uniform* over all fixed composition codes relies on finding the composition which minimizes the lower bound. The optimal composition is given in [87, Eqs. (4.18), (4.19)] for the case where the upper bound on μ'' is applied. In [109], the same composition is used without checking whether it is still the composition which minimizes the lower bound. However (as we see in the next section), for the class of symmetric memoryless channels, the value of the bound is independent of the code composition; therefore, the bound of Valembois and Fossorier [109, Theorem 7] (referred to as the 'VF bound') stays valid. This class of channels includes all memoryless binary-input output-symmetric (MBIOS) channels.

A third improvement in [109] refers to the particular selection of the value of $\rho \geq 0$ which leads to the derivation of Theorem 5.3. In [87], ρ is set to be the value $\tilde{\rho}$ which maximizes the error exponent of the SP67 bound (i.e., the upper bound on the error exponent). This choice emphasizes the similarity between the error exponents of the SP67 bound and the RCB, hence proving that the error exponent of the SP67 bound is tight for all rates above the critical rate of the channel. In order to tighten the bound for finite-length block codes, [109] chooses the value of ρ to be ρ^* which

provides the tightest possible lower bound on the decoding error probability. For rates above the critical rate of the channel, the tightness of the error exponent of the classical SP67 bound implies that $\tilde{\rho}$ tends to ρ^* as the block length tends to infinity. However, for codes of finite block length, this simple observation tightens the bound with almost no penalty in the computational complexity of the resulting bound.

The fourth observation made in [109] refers to the final stage in the derivation of the SP67 bound. In order to get a lower bound on the maximal decoding error probability of an arbitrary block code, the derivation in [87] considers the maximal decoding error probability of a fixed composition subcode of the original code. In [87], a simple lower bound on the size of the largest fixed composition subcode is given; namely, the size of the largest fixed composition subcode is not less than the size of the entire code divided by the number of possible compositions. Since the number of possible compositions is equal to the number of possible ways to divide N symbols into K types, this value is given by $\binom{N+K-1}{K-1}$. To simplify the final expression of the SP67 bound, [87] relies on the inequality $\binom{N+K-1}{K-1} \leq N^K$ which provides a simple upper bound on the number of compositions. Since this expression is polynomial is the block length N, there is no asymptotic loss to the error exponent. However, by using the exact expression for the number of possible compositions, the bound in [109] is tightened for codes of short to moderate block lengths. Applying these four modifications in [109] to the derivation of the SP67 bound yields an improved lower bound on the decoding error probability of block codes transmitted over memoryless channels with finite input alphabets. As mentioned above, these modifications also extend the validity of the new bound to memoryless channels with discrete input and continuous output. However, the requirement of a finite input alphabet still remains, as it is required to apply the bound to arbitrary block codes, and not only to fixed composition codes. Under the assumptions and notation used in Theorem 5.3, the VF bound [109] is given in the following theorem:

Theorem 5.4 [Improvement on the 1967 Sphere-Packing Bound for Discrete Memoryless Channels] [109, Theorem 7]. The *average decoding error probability* satisfies $P_{\mathrm{e}}(N, M, L) \geq \exp\Big(-N E_{\mathrm{VF}}(R, N)\Big)$ where

$$E_{\mathrm{VF}}(R, N) \triangleq \int_{x > \frac{\sqrt{2}}{2}} \Big\{ E_0(\rho_x) - \rho_x \Big(R - O_1\Big(\frac{\ln N}{N}, x\Big)\Big) + O_2\Big(\frac{1}{\sqrt{N}}, x, \rho_x\Big)\Big\}$$

and

$$O_1\Big(\frac{\ln N}{N}, x\Big) \triangleq \frac{\ln 8}{N} + \frac{\ln \binom{N+K-1}{K-1}}{N} - \frac{\ln\big(2 - \frac{1}{x^2}\big)}{N} \tag{5.22}$$

$$O_2\Big(\frac{1}{\sqrt{N}}, x, \rho\Big) \triangleq x\sqrt{\frac{8}{N}\sum_{k=0}^{K-1} q_{k,\rho}\nu_k^{(2)}(\rho)} + \frac{\ln 8}{N} - \frac{\ln\big(2-\frac{1}{x^2}\big)}{N}$$

$$\nu_k^{(1)}(\rho) \triangleq \frac{\displaystyle\sum_{j=0}^{J-1} \beta_{j,k,\rho} \ln\frac{\beta_{j,k,\rho}}{P(j|k)}}{\displaystyle\sum_{j=0}^{J-1}\beta_{j,k,\rho}}$$

$$\nu_k^{(2)}(\rho) \triangleq \frac{\displaystyle\sum_{j=0}^{J-1} \beta_{j,k,\rho} \ln^2\frac{\beta_{j,k,\rho}}{P(j|k)}}{\displaystyle\sum_{j=0}^{J-1}\beta_{j,k,\rho}} - \big[\nu_k^{(1)}(\rho)\big]^2$$

$$\beta_{j,k,\rho} \triangleq P(j|k)^{\frac{1}{1+\rho}} \cdot \left(\sum_{k'=0}^{K-1} q_{k',\rho} P(j|k')^{\frac{1}{1+\rho}}\right)^{\rho}$$

where $\mathbf{q}_\rho \triangleq (q_{1,\rho}, \ldots, q_{K,\rho})$ designates the input distribution which maximizes $E_0(\rho, \mathbf{q})$ in (5.19), and the parameter $\rho = \rho_x$ is determined by solving the equation

$$R - O_1\Big(\frac{\ln N}{N}, x\Big) = -\frac{1}{\rho}\sum_{k=0}^{K-1} q_{k,\rho}\nu_k^{(1)}(\rho) + \frac{x}{\rho}\sqrt{\frac{2}{N}\sum_{k=0}^{K-1} q_{k,\rho}\nu_k^{(2)}(\rho)}.$$

For a more detailed review of the improvements suggested in [109], the reader is referred to [79, Section 5.4].

Remark 5.1 The rate loss as a result of the expurgation of the code by removing half of the codewords with the largest error probability was ignored in [109]. The term $\frac{\ln 4}{N}$, as it appears in the term $O_1(\frac{\ln N}{N}, x)$ of [109, Theorem 7], should be therefore replaced by $\frac{\ln 8}{N}$ (see (5.22)).

5.3 An Improved Sphere-Packing Bound for Symmetric Memoryless Channels

In this section, we derive an improved lower bound on the decoding error probability which utilizes the sphere-packing bounding technique. This new bound is valid for *symmetric* memoryless channels with a finite input alphabet, and is referred to as an improved sphere-packing (ISP) bound. Note that the symmetry of the channel is crucial for the derivation of the ISP bound in this section, which stays in contrast to the SP67 and VF bounds where channel symmetry is not required. We begin with

some necessary definitions and basic properties of symmetric memoryless channels which are used in this section for the derivation of the ISP bound.

5.3.1 Symmetric Memoryless Channels

Definition 5.1 A bijective mapping $g : \mathcal{J} \to \mathcal{J}$ where $\mathcal{J} \subseteq \mathbb{R}^d$ is said to be *unitary* if for any integrable generalized function $f : \mathcal{J} \to \mathbb{R}$

$$\int_{\mathcal{J}} f(x)\mathrm{d}x = \int_{\mathcal{J}} f(g(x))\mathrm{d}x \tag{5.23}$$

where by generalized function we mean a function which may contain a countable number of shifted Dirac delta functions. If the projection of $\mathcal{J}$ over some of the d dimensions is countable, the integration over these dimensions is turned into a sum.

Remark 5.2 The following properties also hold:

1. If g is a unitary mapping so is its inverse g^{-1}.
2. If $\mathcal{J}$ is a countable set, then $g : \mathcal{J} \to \mathcal{J}$ is unitary if and only if g is bijective.
3. Let $\mathcal{J}$ be an open set and $g : \mathcal{J} \to \mathcal{J}$ be a bijective function. Denote

$$g(x_1, \ldots, x_d) \triangleq \big(g_1(x_1, \ldots, x_d), \ldots, g_d(x_1, \ldots, x_d)\big)$$

and assume that the partial derivatives $\frac{\partial g_i}{\partial x_j}$ exist for all $i, j \in \{1, 2, \ldots, d\}$. Then g is unitary if and only if the Jacobian satisfies $|J(\mathbf{x})| = 1$ for all $\mathbf{x} \in \mathcal{J}$.

Proof: The first property follows from (5.23) and by defining $\widetilde{f}(x) \triangleq f\big(g^{-1}(x)\big)$; this gives

$$\int_{\mathcal{J}} f\big(g^{-1}(x)\big)\mathrm{d}x = \int_{\mathcal{J}} \widetilde{f}(x)\mathrm{d}x = \int_{\mathcal{J}} \widetilde{f}\big(g(x)\big)\mathrm{d}x = \int_{\mathcal{J}} f\big((g^{-1} \circ g)(x)\big)\mathrm{d}x = \int_{\mathcal{J}} f(x)\mathrm{d}x\,.$$

The second property follows from the fact that for countable sets, the integral is turned into a sum, and the equality

$$\sum_{j \in \mathcal{J}} f(j) = \sum_{j \in \mathcal{J}} f\big(g(j)\big)$$

holds by changing the order of summation. Finally, the third property is proved by a transform of the integrator on the LHS of (5.23) from $\mathbf{x} = (x_1, \ldots, x_d)$ to $g(\mathbf{x})$. ■

We are now ready to define K-ary input symmetric channels. The symmetry properties of these channels are later exploited to improve the tightness of the sphere-packing bounding technique and derive the ISP lower bound on the average decoding error probability of block codes transmitted over these channels.

Definition 5.2 [**Symmetric Memoryless Channels**] A memoryless channel with input alphabet $\mathcal{K} = \{0, 1, \ldots, K-1\}$, output alphabet $\mathcal{J} \subseteq \mathbb{R}^d$ (where $K, d \in \mathbb{N}$) and transition probability (or density if $\mathcal{J}$ non-countable) function $P(\cdot|\cdot)$ is said to be *symmetric* if there exists a set of bijective and unitary mappings $\{g_k\}_{k=0}^{K-1}$ where $g_k : \mathcal{J} \to \mathcal{J}$ for all $k \in \mathcal{K}$ such that

$$\forall \mathbf{y} \in \mathcal{J},\, k \in \mathcal{K} \quad P(\mathbf{y}|0) = P\big(g_k(\mathbf{y})|k\big) \tag{5.24}$$

and

$$\forall k_1, k_2 \in \mathcal{K} \quad g_{k_1}^{-1} \circ g_{k_2} = g_{(k_2 - k_1) \bmod K}\,. \tag{5.25}$$

Remark 5.3 From (5.24), the mapping g_0 is the identity mapping. Assigning $k_1 = k$ and $k_2 = 0$ in (5.25) gives

$$\forall k \in \mathcal{K} \quad g_k^{-1} = g_{(-k) \bmod K} = g_{K-k}\,. \tag{5.26}$$

The class of symmetric memoryless channels, as given in Definition 5.2, is quite large. In particular, it contains the class of memoryless binary-input output-symmetric (MBIOS) channels. To show this, we employ the following proposition which follows from the discussion in [74, Section 4.1.4]:

Proposition 5.1 An MBIOS channel can be equivalently represented as a (time-varying) binary symmetric channel (BSC) whose crossover probability for each output symbol is an i.i.d. random variable which is independent of the channel input, and observed by the receiver. This crossover probability is given by $p = \frac{1}{1+\exp(|L|)}$ where $L = L(y)$ denotes the log-likelihood ratio which corresponds to the channel output y.

We now apply Proposition 5.1 to show that any MBIOS channel is a symmetric memoryless channel, according to Definition 5.2.

Corollary 5.1 An arbitrary MBIOS channel, can be equivalently represented as a symmetric memoryless channel.

Proof: Let us consider an MBIOS channel $\mathfrak{C}$. Applying Proposition 5.1, it can be equivalently represented by a channel $\mathfrak{C}'$ whose output alphabet is $\mathcal{J} = \{0, 1\} \times [0, 1]$; here, the first term of the output refers to the BSC output and the second term is the associated crossover probability. We now show that this equivalent channel is a symmetric memoryless channel. To this end, it suffices to find a unitary mapping $g_1 : \mathcal{J} \to \mathcal{J}$ such that

$$\forall \mathbf{y} \in \mathcal{J} \quad P(\mathbf{y}|0) = P\big(g_1(\mathbf{y})|1\big) \tag{5.27}$$

and $g_1^{-1} = g_1$ (i.e., g_1 is equal to its inverse).

For the channel $\mathfrak{C}'$, the conditional probability distribution (or density) function of the output $\mathbf{y} = (m, p)$ (where $m \in \{0, 1\}$ and $p \in [0, 1]$) given that $i \in \{0, 1\}$ is transmitted, is given by

$$P(\mathbf{y}|i) = \begin{cases} \widetilde{P}(p) \cdot (1-p) & \text{if } i = m \\ \widetilde{P}(p) \cdot p & \text{if } i = \overline{m} \end{cases} \tag{5.28}$$

where $\widetilde{P}$ is a distribution (or density) over $[0, 1]$ and $\overline{m}$ designates the logical not of m. From (5.28), we get that the mapping $g_1(m, p) = \big(\overline{m}, p\big)$ satisfies (5.27). Additionally, $g_1^{-1} = g_1$ since $\overline{\overline{m}} = m$. Therefore, the proof is completed by showing that g_1 is a unitary mapping. For any (generalized) function $f : \mathcal{J} \to \mathbb{R}$ we have

$$\begin{aligned} \int_{\mathcal{J}} f(\mathbf{x}) \mathrm{d}\mathbf{x} &\triangleq \sum_{m=0}^{1} \int_0^1 f(m, p) \mathrm{d}p \\ &= \sum_{m=0}^{1} \int_0^1 f(\bar{m}, p) \mathrm{d}p \\ &= \int_{\mathcal{J}} f\big(g_1(\mathbf{x})\big) \mathrm{d}\mathbf{x} \end{aligned}$$

where the second equality holds by changing the order of summation; hence g_1 is a unitary function. ■

Remark 5.4 Proposition 5.1 forms a special case of a proposition given in [113, Appendix I]. Using the proposition in [113, Appendix I], which refers to M-ary input channels, it can be shown in a similar way that all M-ary input symmetric output channels, as defined in [113], can be equivalently represented as symmetric memoryless channels.

Coherently detected M-ary PSK modulated signals transmitted over a fully interleaved fading channel, followed by an additive white Gaussian noise, form another example of a symmetric memoryless channel. In this case, $\mathcal{J} = \mathbb{R}^2$ and the mapping g_k for $k = 0, \ldots, M-1$ forms a clockwise rotation by $\frac{2\pi k}{M}$ (i.e., $g_k(\mathbf{y}) = \exp\big(\frac{2j\pi k}{M}\big)\, \mathbf{y}$). Note that the determinant of the Jacobian of these rotation mappings is equal in absolute value to 1.

5.3.2 Derivation of an Improved Sphere-Packing Bound for Symmetric Memoryless Channels

In this section, we derive an improved sphere-packing lower bound on the decoding error probability of block codes transmitted over symmetric memoryless channels.

To keep the notation simple, we derive the bound under the assumption that the communication takes place over a symmetric DMC. However, the derivation of the bound is justified later for the general class of symmetric memoryless channels with discrete or continuous output alphabets. Some remarks are given at the end of the derivation.

Though there is a certain parallelism to the derivation of the SP67 bound in [87, Part 1], our analysis for symmetric memoryless channels deviates considerably from the derivation of this classical bound. The improvements suggested in [109] are also incorporated into the derivation of the bound. We show that for symmetric memoryless channels, the derivation of the sphere-packing bound can be modified so that the intermediate step of bounding the maximal error probability for fixed composition codes can be skipped, and one can directly consider the *average* error probability of an *arbitrary* block code. To this end, the first step of the derivation in [87] (see Theorem 5.1 here) is modified so that instead of bounding the error probability when a single pair of probability assignments is considered, we consider the average error probability over M pairs of probability assignments.

Average Decoding Error Probability for M Pairs of Probability Assignments

We start the analysis by considering the average decoding error probability over M pairs of probability assignments, denoted $\{P_1^m, P_2^m\}_{m=1}^M$, where it is assumed that the index m of the pair is chosen uniformly at random from the set $\{1, \ldots, M\}$ and is known to the decoder. Denote the observation by $\mathbf{y}$ and the observation space by $\mathcal{Y}$. For simplicity, we assume that $\mathcal{Y}$ is a finite set. Following the notation in [87], we define the LLR for the m^{th} pair of probability assignments as

$$D^m(\mathbf{y}) \triangleq \ln\left(\frac{P_2^m(\mathbf{y})}{P_1^m(\mathbf{y})}\right) \tag{5.29}$$

and the probability distribution

$$Q_s^m(\mathbf{y}) \triangleq \frac{P_1^m(\mathbf{y})^{1-s}\, P_2^m(\mathbf{y})^s}{\sum_{\mathbf{y}'} P_1^m(\mathbf{y}')^{1-s}\, P_2^m(\mathbf{y}')^s}\,, \quad 0 \le s \le 1\,. \tag{5.30}$$

For the m^{th} pair, we also define the function μ^m as

$$\mu^m(s) \triangleq \ln\left(\sum_{\mathbf{y}} P_1^m(\mathbf{y})^{1-s}\, P_2^m(\mathbf{y})^s\right)\,, \quad 0 \le s \le 1\,. \tag{5.31}$$

Let us assume that μ^m and its first and second derivatives w.r.t. s are independent of the value of m, and therefore we can define $\mu \triangleq \mu^1 = \mu^2 = \ldots = \mu^M$.

Remark 5.5 Note that in this setting, the requirement that μ^m is independent of m inherently yields that all its derivatives are also independent of m. However, in the continuation, we will let P_2^m be a function of s and differentiate μ^m w.r.t. s while holding P_2^m fixed. In this setting, we will show that for the specific selection of P_1^m and P_2^m which are used to derive the new lower bound on the average block error probability, if the communication takes place over a symmetric memoryless channel then μ^m and its first two derivatives w.r.t. s are independent of m. Also note that the fact that μ^m is independent of m does not imply that P_k^m is independent of m.

Based on the assumption above, it can be easily verified (in parallel to (5.9)–(5.12)) that for all $m \in \{1, \ldots, M\}$

$$\mu'(s) = \big(\mu^m\big)'(s) = \mathbb{E}_{Q_s^m}\big(D^m(\mathbf{y})\big) \tag{5.32}$$

$$\mu''(s) = \big(\mu^m\big)''(s) = \mathrm{Var}_{Q_s^m}\big(D^m(\mathbf{y})\big) \tag{5.33}$$

$$P_1^m(\mathbf{y}) = \exp\big(\mu(s) - sD^m(\mathbf{y})\big)\, Q_s^m(\mathbf{y}) \tag{5.34}$$

$$P_2^m(\mathbf{y}) = \exp\big(\mu(s) + (1-s)D^m(\mathbf{y})\big)\, Q_s^m(\mathbf{y}) \tag{5.35}$$

where $\mathbb{E}_Q$ and Var_Q stand, respectively, for the statistical expectation and variance w.r.t. a probability distribution Q. For the m^{th} code book, we define the set of typical output vectors as

$$\mathcal{Y}_s^{m,x} \triangleq \Big\{\mathbf{y} \in \mathcal{Y} : |D^m(\mathbf{y}) - \mu'(s)| \le x\sqrt{2\mu''(s)}\Big\}, \quad x > 0. \tag{5.36}$$

In the original derivation of the SP67 bound in [87] (see (5.13) here), the parameter x was set to one; similarly to [109], this parameter is introduced in (5.36) in order to tighten the bound for finite-length block codes. However, in both [87] and [109], only one pair of probability assignments was considered. By applying Chebychev's inequality to (5.36), and relying on the equalities in (5.32) and (5.33), we get that for all $m \in \{1, \ldots, M\}$

$$\sum_{\mathbf{y} \in \mathcal{Y}_s^{m,x}} Q_s^m(\mathbf{y}) > 1 - \frac{1}{2x^2} \tag{5.37}$$

where this result is meaningful only for $x > \frac{\sqrt{2}}{2}$.

Let $\mathcal{Y}_1^m$ and $\mathcal{Y}_2^m$ be the decoding regions of P_1^m and P_2^m, respectively. Since the index m is known to the decoder, P_1^m is decoded only against P_2^m; hence, $\mathcal{Y}_1^m$ and $\mathcal{Y}_2^m$ form a partition of the observation space $\mathcal{Y}$. We now derive a lower bound on the conditional error probability given that the correct hypothesis is the first probability assignment and the m^{th} pair was selected. Similarly to (5.14), we get the following lower bound from (5.34) and (5.36):

$$P_{\text{e},1}^m \ge \exp\Big(\mu(s) - s\mu'(s) - s\,x\,\sqrt{2\mu''(s)}\Big) \sum_{\mathbf{y} \in \mathcal{Y}_2^m \bigcap \mathcal{Y}_s^{m,x}} Q_s^m(\mathbf{y}). \tag{5.38}$$

Following the same steps w.r.t. the conditional error probability of P_2^m and applying (5.35), gives

$$P_{\mathrm{e},2}^m \geq \exp\Big(\mu(s) + (1-s)\mu'(s) - (1-s)\,x\,\sqrt{2\mu''(s)}\Big) \sum_{\mathbf{y}\in\mathcal{Y}_1^m \bigcap \mathcal{Y}_s^{m,x}} Q_s^m(\mathbf{y}). \tag{5.39}$$

Averaging (5.38) and (5.39) over m gives that for all $s \in (0,1)$

$$\begin{aligned} P_{\mathrm{e},1}^{\mathrm{avg}} &\triangleq \frac{1}{M}\sum_{m=1}^{M} P_{\mathrm{e},1}^m \\ &\geq \exp\Big(\mu(s) - s\mu'(s) - s\,x\,\sqrt{2\mu''(s)}\Big)\,\frac{1}{M}\sum_{m=1}^{M}\sum_{\mathbf{y}\in\mathcal{Y}_2^m \bigcap \mathcal{Y}_s^{m,x}} Q_s^m(\mathbf{y}) \end{aligned} \tag{5.40}$$

and

$$\begin{aligned} P_{\mathrm{e},2}^{\mathrm{avg}} &\triangleq \frac{1}{M}\sum_{m=1}^{M} P_{\mathrm{e},2}^m \\ &\geq \exp\Big(\mu(s) + (1-s)\mu'(s) - (1-s)x\sqrt{2\mu''(s)}\Big)\frac{1}{M}\sum_{m=1}^{M}\sum_{\mathbf{y}\in\mathcal{Y}_1^m \bigcap \mathcal{Y}_s^{m,x}} Q_s^m(\mathbf{y}) \end{aligned} \tag{5.41}$$

where $P_{\mathrm{e},1}^{\mathrm{avg}}$ and $P_{\mathrm{e},2}^{\mathrm{avg}}$ refer to the average error probabilities given that the first or second hypotheses, respectively, of a given pair are correct where this pair is chosen uniformly at random among the M possible pairs of hypotheses. Since for all m, the sets $\mathcal{Y}_1^m$ and $\mathcal{Y}_2^m$ form a partition of the set of output vectors $\mathcal{Y}$, then

$$\frac{1}{M}\sum_{m=1}^{M}\sum_{\mathbf{y}\in\mathcal{Y}_1^m \bigcap \mathcal{Y}_s^{m,x}} Q_s^m(\mathbf{y}) + \frac{1}{M}\sum_{m=1}^{M}\sum_{\mathbf{y}\in\mathcal{Y}_2^m \bigcap \mathcal{Y}_s^{m,x}} Q_s^m(\mathbf{y}) = \frac{1}{M}\sum_{m=1}^{M}\sum_{\mathbf{y}\in\mathcal{Y}_s^{m,x}} Q_s^m(\mathbf{y}) > 1 - \frac{1}{2x^2}$$

where the last transition follows from (5.37) and is meaningful for $x > \frac{\sqrt{2}}{2}$. Hence, at least one of the terms in the LHS of the above equality is necessarily greater than $\frac{1}{2}\left(1 - \frac{1}{2x^2}\right)$. Combining this result with (5.40) and (5.41), we get that for every $s \in (0,1)$

$$P_{\mathrm{e},1}^{\mathrm{avg}} > \left(\frac{1}{2} - \frac{1}{4x^2}\right) \exp\Big(\mu(s) - s\mu'(s) - s\,x\,\sqrt{2\mu''(s)}\Big) \tag{5.42}$$

or

$$P_{\mathrm{e},2}^{\mathrm{avg}} > \left(\frac{1}{2} - \frac{1}{4x^2}\right) \exp\Big(\mu(s) + (1-s)\mu'(s) - (1-s)\,x\,\sqrt{2\mu''(s)}\Big). \tag{5.43}$$

The two inequalities above provide a lower bound on the average decoding error probability over M pairs of probability assignments.

We now turn to consider a block code which is transmitted over a symmetric DMC. Similarly to the derivation of the SP67 bound in [87], we use the lower bound derived

in this section to relate the decoding error probability when a given codeword is transmitted to the size of the decision region associated with this codeword. However, the bound above allows us to directly consider the average block error probability; this is in contrast to the derivation in [87] which first considered the maximal block error probability of the code and then used an argument based on expurgating half of the bad codewords in order to obtain a lower bound on the average error probability of the original code (where the code rate is asymptotically not affected as a result of this expurgation). Additionally, we show that when the transmission takes place over a memoryless symmetric channel, one can consider directly an arbitrary block code instead of starting the analysis by referring to fixed composition codes as in [87, Part 1] and [109].

Lower Bound on the Decoding Error Probability of General Block Codes

We now consider a block code $\mathcal{C}$ of length N with M codewords, denoted by $\{\mathbf{x_m}\}_{m=1}^{M}$; assume that the transmission takes place over a symmetric DMC with transition probabilities $P(j|k)$, where $k \in \mathcal{K} = \{0, \ldots, K-1\}$ and $j \in \mathcal{J} = \{0, \ldots, J-1\}$ designate the channel input and output alphabets, respectively. In this section, we derive a lower bound on the average block error probability of the code $\mathcal{C}$ for an arbitrary list decoder where the size of the list is limited to L. Let f_N be a probability measure defined over the set of length-N sequences of the channel output, and which can be factorized as in (5.16). We define M pairs of probability measures $\{P_1^m, P_2^m\}$ by

$$P_1^m(\mathbf{y}) \triangleq \Pr(\mathbf{y}|\mathbf{x_m}), \quad P_2^m(\mathbf{y}) \triangleq f_N(\mathbf{y}), \quad m \in \{1, 2, \ldots, M\} \tag{5.44}$$

where $\mathbf{x_m}$ is the m^{th} codeword of the code $\mathcal{C}$. Combining (5.31) and (5.44), the function μ^m takes the form

$$\mu^m(s) = \ln\left(\sum_{\mathbf{y}} \Pr(\mathbf{y}|\mathbf{x_m})^{1-s} f_N(\mathbf{y})^s\right), \quad 0 < s < 1. \tag{5.45}$$

Let us denote by q_k^m the fraction of appearances of the letter k in the codeword $\mathbf{x_m}$. By assumption, the communication channel is memoryless and the function f_N is a probability measure which is factorized according to (5.16). Hence, for every $m \in \{1, 2, \ldots, M\}$, the function μ^m in (5.45) is expressible in the form

$$\mu^m(s) = N \sum_{k=0}^{K-1} q_k^m \mu_k(s) \tag{5.46}$$

where

$$\mu_k(s) \triangleq \ln\left(\sum_{j=0}^{J-1} P(j|k)^{1-s} f(j)^s\right), \quad 0 < s < 1. \tag{5.47}$$

In order to validate the statement which assures that at least one of the inequalities in (5.42) and (5.43) is satisfied, it is required to verify in this case that the function μ^m and its first and second derivatives w.r.t. s are independent of the index m. From (5.46), since $\sum_{k=0}^{K-1} q_k^m = 1$ for every $m \in \{1, \ldots, M\}$, it suffices to show that μ_k and its first and second derivatives are independent of the input symbol k. To this end, for every $s \in (0,1)$, we choose the function f to be f_s, as given in [87, Eqs. (4.18)–(4.20)]. Namely, for $0 < s < 1$, let $\mathbf{q_s} = \{q_{0,s}, \ldots, q_{K-1,s}\}$ satisfy the inequalities

$$\sum_{j=0}^{J-1} P(j|k)^{1-s} \alpha_{j,s}^{\frac{s}{1-s}} \geq \sum_{j=0}^{J-1} \alpha_{j,s}^{\frac{1}{1-s}}; \quad \forall k \in \mathcal{K} \tag{5.48}$$

where

$$\alpha_{j,s} \triangleq \sum_{k'=0}^{K-1} q_{k',s} P(j|k')^{1-s}. \tag{5.49}$$

The function $f = f_s$ is given by

$$f_s(j) = \frac{\alpha_{j,s}^{\frac{1}{1-s}}}{\sum_{j'=0}^{J-1} \alpha_{j',s}^{\frac{1}{1-s}}}, \quad j \in \{0, \ldots, J-1\}. \tag{5.50}$$

Note that the input distribution $\mathbf{q_s}$ is *independent of the code* $\mathcal{C}$, as it only depends on the channel statistics. It should be also noted that P_1^m and P_2^m are in general allowed to depend on the parameter s, though the differentiation of the function μ^m w.r.t. s is performed while holding P_1^m and P_2^m fixed. The following lemma shows that for symmetric channels, the function f_s in (5.50) yields that μ_k and its first and second derivatives w.r.t. s (while holding f_s fixed) are independent of the input symbol k.

Lemma 5.1 Let $P(\cdot|\cdot)$ designate the transition probability function of a symmetric DMC with input alphabet $\mathcal{K} = \{0, \ldots, K-1\}$ and output alphabet $\mathcal{J} = \{0, \ldots, J-1\}$, and let μ_k be defined as in (5.47), where $f = f_s$ is given in (5.50). Then, the following properties hold for all $s \in (0,1)$

$$\mu_0(s) = \mu_1(s) = \ldots = \mu_{K-1}(s) = -(1-s) E_0\left(\frac{s}{1-s}\right) \tag{5.51}$$

$$\mu_0'(s) = \mu_1'(s) = \ldots = \mu_{K-1}'(s) \tag{5.52}$$

$$\mu_0''(s) = \mu_1''(s) = \ldots = \mu_{K-1}''(s) \tag{5.53}$$

where E_0 is introduced in (5.19) and the differentiation in (5.52) and (5.53) is performed w.r.t s while holding f_s fixed.

Proof: The proof of this lemma is quite technical and is given in Appendix 5.A. ■

Remark 5.6 Since the differentiation of the function μ_k w.r.t. s is performed while holding $f = f_s$ fixed, then the independence of the function μ_k in the parameter k, as stated in (5.51), does not necessarily imply the independence of the first and second derivatives of μ_k as in (5.52) and (5.53); in order to prove Lemma 5.1 (see Appendix 5.A), we rely on the symmetry of the memoryless channel. The function μ_0 in (5.4) and its derivatives are calculated in Appendix 5.B for some symmetric memoryless channels, and these results are later used for the numerical calculations of the sphere-packing bounds in Section 5.5.

By (5.46) and Lemma 5.1, we get that the function μ^m and its first and second derivatives w.r.t. s are independent of the index m (where this property also follows since $\sum_{k=0}^{K-1} q_k^m = 1$, irrespectively of m).

Let $\mathcal{Y}_m$ be the decision region of the codeword $\mathbf{x_m}$. By associating $\mathcal{Y}_m$ and $\mathcal{Y}_m^{\mathrm{c}}$ with the two decision regions for the probability measures P_1^m and P_2^m, respectively, we get from (5.44)

$$P_{\mathrm{e},1}^m = \sum_{\mathbf{y}\in\mathcal{Y}_m^{\mathrm{c}}} P_1^m(\mathbf{y}) = \sum_{\mathbf{y}\in\mathcal{Y}_m^{\mathrm{c}}} \Pr(\mathbf{y}|\mathbf{x_m}) \triangleq P_{\mathrm{e},m}$$

and

$$P_{\mathrm{e},2}^m = \sum_{\mathbf{y}\in\mathcal{Y}_m} P_2^m(\mathbf{y}) = \sum_{\mathbf{y}\in\mathcal{Y}_m} f_N(\mathbf{y}) = F(\mathcal{Y}_m)$$

where $P_{\mathrm{e},m}$ is the decoding error probability of the code $\mathcal{C}$ when the codeword $\mathbf{x_m}$ is transmitted, and $F(\mathcal{Y}_m)$ is a measure for the size of the decoding region $\mathcal{Y}_m$ as defined in (5.17). Substituting the two equalities above in (5.42) and (5.43) gives that for all $s \in (0,1)$

$$\frac{1}{M}\sum_{m=1}^{M} P_{\mathrm{e},m} = P_{\mathrm{e},1}^{\mathrm{avg}} > \left(\frac{1}{2} - \frac{1}{4x^2}\right) \exp\Big(\mu(s) - s\mu'(s) - s\,x\,\sqrt{2\mu''(s)}\Big) \tag{5.54}$$

or

$$\frac{1}{M}\sum_{m=1}^{M} F_s(\mathcal{Y}_m) = P_{\mathrm{e},2}^{\mathrm{avg}} > \left(\frac{1}{2} - \frac{1}{4x^2}\right) \exp\Big(\mu(s) + (1-s)\mu'(s) - (1-s)\,x\,\sqrt{2\mu''(s)}\Big) \tag{5.55}$$

where $x > \frac{\sqrt{2}}{2}$ and $F_s(\mathcal{Y}_m) \triangleq \sum_{\mathbf{y}\in\mathcal{Y}_m} f_{N,s}(\mathbf{y})$. Similarly to [87], we relate $\sum_{m=1}^{M} F_s(\mathcal{Y}_m)$ to the number of codewords M and to the size of the decoding list which is limited to L. First, for all $0 \le s \le 1$

$$\sum_{m=1}^{M} F_s(\mathcal{Y}_m) = \sum_{m=1}^{M} \sum_{\mathbf{y}\in\mathcal{Y}_m} f_{N,s}(\mathbf{y}) \le L$$

where the last inequality holds since each $\mathbf{y} \in \mathcal{J}^N$ is included in at most L subsets $\{\mathcal{Y}_m\}_{m=1}^M$ and $\sum_{\mathbf{y}} f_{N,s}(\mathbf{y}) = 1$. Hence, the LHS of (5.55) is upper bounded by $\frac{L}{M}$ for all $0 \leq s \leq 1$. Additionally, the LHS of (5.54) is equal by definition to the average block error probability P_{e} of the code $\mathcal{C}$. Therefore, (5.54) and (5.55) can be rewritten as

$$P_{\mathrm{e}} > \left(\frac{1}{2} - \frac{1}{4x^2}\right) \exp\Big(\mu(s) - s\mu'(s) - s\,x\,\sqrt{2\mu''(s)}\Big) \tag{5.56}$$

or

$$\frac{L}{M} > \left(\frac{1}{2} - \frac{1}{4x^2}\right) \exp\Big(\mu(s) + (1-s)\mu'(s) - (1-s)\,x\,\sqrt{2\mu''(s)}\Big). \tag{5.57}$$

Applying (5.46) and Lemma 5.1 to (5.56) and (5.57) gives that for all $s \in (0,1)$

$$P_{\mathrm{e}} > \left(\frac{1}{2} - \frac{1}{4x^2}\right) \exp\left\{N\left(\mu_0(s,f_s) - s\mu_0'(s,f_s) - s\,x\,\sqrt{\frac{2\mu_0''(s,f_s)}{N}}\right)\right\} \tag{5.58}$$

or

$$\frac{L}{M} > \left(\frac{1}{2} - \frac{1}{4x^2}\right) \exp\left\{N\left(\mu_0(s,f_s) + (1-s)\mu_0'(s,f_s) - (1-s)x\sqrt{\frac{2\mu_0''(s,f_s)}{N}}\right)\right\}. \tag{5.59}$$

A lower bound on the average block error probability can be obtained from (5.58) by substituting any value of $s \in (0,1)$ for which the inequality in (5.59) does not hold. In particular we choose a value $s = s_x$ such that the inequality in (5.59) is replaced by an equality, i.e.,

$$\begin{aligned}\frac{L}{M} &= \exp(-NR) \\ &= \left(\frac{1}{2} - \frac{1}{4x^2}\right) \exp\left\{N\Bigg(\mu_0(s_x,f_{s_x}) + (1-s_x)\,\mu_0'(s_x,f_{s_x}) \right. \\ &\qquad \left. -(1-s_x)\,x\,\sqrt{\frac{2\mu_0''(s_x,f_{s_x})}{N}}\Bigg)\right\}\end{aligned} \tag{5.60}$$

where $R \triangleq \frac{\ln\left(\frac{M}{L}\right)}{N}$ designates the code rate in nats per channel use. Note that the existence of a solution $s = s_x$ to (5.60) can be demonstrated in a similar way to the arguments in [87, Eqs. (4.28)–(4.35)] for the non-trivial case where the sphere-packing bound does not reduce to the trivial inequality $P_{\mathrm{e}} \geq 0$. This particular value of s is chosen since for a large enough value of N, the RHS of (5.58) is monotonically decreasing while the RHS of (5.59) is monotonically increasing for $s \in (0,1)$; thus, this choice is optimal for large enough N. The choice of $s = s_x$ also allows to get a simpler representation of the bound on the average block error probability. Rearranging (5.60) gives

$$\mu_0'(s_x,f_{s_x}) = -\frac{1}{1-s_x}\left[R + \mu_0(s_x,f_{s_x}) + \frac{1}{N}\,\ln\left(\frac{1}{2} - \frac{1}{4x^2}\right)\right] + x\sqrt{\frac{2\mu_0''(s_x,f_{s_x})}{N}}.$$

Substituting $s = s_x$ and the last equality into (5.58) yields that

$$P_{\mathrm{e}} \;>\; \exp\left\{ N\left(\frac{\mu_0(s_x, f_{s_x})}{1-s_x} + \frac{s_x}{1-s_x}\left(R + \frac{1}{N}\ln\left(\frac{1}{2} - \frac{1}{4x^2}\right)\right) \right.\right. \\ \left.\left. -s_x\, x \sqrt{\frac{8\mu_0''(s_x, f_{s_x})}{N}} + \frac{1}{N}\ln\left(\frac{1}{2} - \frac{1}{4x^2}\right)\right)\right\}.$$

By applying (5.51) and defining $\rho_x \triangleq \frac{s_x}{1-s_x}$ we get

$$P_{\mathrm{e}} \;>\; \exp\left\{ -N\left(E_0(\rho_x) - \rho_x\left[R + \frac{1}{N}\ln\left(\frac{1}{2} - \frac{1}{4x^2}\right)\right] \right.\right. \\ \left.\left. +s_x\, x \sqrt{\frac{8\mu_0''(s_x, f_{s_x})}{N}} - \frac{1}{N}\ln\left(\frac{1}{2} - \frac{1}{4x^2}\right)\right)\right\}.$$

Note that the above lower bound on the average decoding error probability holds for an arbitrary block code of length N and rate R. The selection of ρ_x is similar to [109]. Finally, we optimize over the parameter $x \in (\frac{\sqrt{2}}{2}, \infty)$ in order to get the tightest lower bound of this form.

The derivation above only relies on the fact that the channel is memoryless and symmetric, but does not rely on the fact that the output alphabet is discrete. As mentioned in Section 5.2.2, the original derivation of the SP67 bound in [87] relies on the fact that the input and output alphabets are finite in order to upper bound $\mu''(s)$ by $\left(\frac{1}{s}\ln\left(\frac{e}{\sqrt{P_{\min}}}\right)\right)^2$ where $P_{\min}$ designates the smallest non-zero transition probability of the channel. This requirement was relaxed in [109] to the requirement that only the input alphabet is finite; to this end, the second derivative of the function μ is calculated, thus the above upper bound on this second derivative is replaced by its exact value. The validity of the derivation for symmetric continuous-output channels is considered in the continuation (see Remark 5.9). This leads to the following theorem, which provides an improved sphere-packing lower bound on the decoding error probability of block codes transmitted over symmetric memoryless channels.

Theorem 5.5 [An Improved Sphere-Packing (ISP) Bound for Symmetric Memoryless Channels] Let $\mathcal{C}$ be an arbitrary block code consisting of M codewords, each of length N. Assume that $\mathcal{C}$ is transmitted over a memoryless symmetric channel which is specified by the transition probabilities (or densities) $P(j|k)$ where $k \in \mathcal{K} = \{0, \ldots, K-1\}$ and $j \in \mathcal{J} \subseteq \mathbb{R}^d$ designate the channel input and output alphabets, respectively. Assume an arbitrary list decoder where the size of the list is limited to L. Then, the *average decoding error probability* satisfies

$$P_{\mathrm{e}}(N, M, L) \geq \exp\Big\{ -N E_{\mathrm{ISP}}(R, N) \Big\}$$

where

$$E_{\text{ISP}}(R,N) \triangleq \inf_{x>\frac{\sqrt{2}}{2}} \Big\{ E_0(\rho_x) - \rho_x \Big(R - O_1\Big(\frac{1}{N}, x\Big)\Big) + O_2\Big(\frac{1}{\sqrt{N}}, x, \rho_x\Big)\Big\} \tag{5.61}$$

the function E_0 is introduced in (5.19), $R = \frac{1}{N}\ln\Big(\frac{M}{L}\Big)$, and

$$O_1\Big(\frac{1}{N}, x\Big) \triangleq -\frac{1}{N}\ln\left(\frac{1}{2} - \frac{1}{4x^2}\right) \tag{5.62}$$

$$O_2\Big(\frac{1}{\sqrt{N}}, x, \rho\Big) \triangleq s(\rho)\, x \sqrt{\frac{8}{N}\mu_0''\big(s(\rho), f_{s(\rho)}\big)} - \frac{1}{N}\ln\left(\frac{1}{2} - \frac{1}{4x^2}\right). \tag{5.63}$$

Here, $s(\rho) \triangleq \frac{\rho}{1+\rho}$, and the non-negative parameter $\rho = \rho_x$ on the RHS of (5.61) is determined by solving the equation

$$R - O_1\Big(\frac{1}{N}, x\Big) = -\mu_0\big(s(\rho), f_{s(\rho)}\big) - \big(1-s(\rho)\big)\mu_0'\big(s(\rho), f_{s(\rho)}\big) + \big(1-s(\rho)\big)\, x \sqrt{\frac{2\mu_0''\big(s(\rho), f_{s(\rho)}\big)}{N}} \tag{5.64}$$

and the functions $\mu_0(s,f)$ and f_s are defined in (5.47) and (5.50), respectively.

Remark 5.7 The requirement that the communication channel is symmetric is crucial to the derivation of the ISP bound. One of the new concepts introduced here is the use of the channel symmetry to show that the function μ^m and its first and second derivatives w.r.t. s are independent of the codeword composition. This enables to tighten the VF bound in [109] by skipping the intermediate step which is related to fixed composition codes. Another new concept is a direct consideration of the *average decoding error probability* of the code rather than considering the maximal block error probability and expurgating the code. This is due to the consideration of M pairs of probability distributions in the first step of the derivation. Note that the bound on the average block error probability of M probability assignment pairs requires that μ^m and its first and second derivatives are independent of the index m; this property holds due to the symmetry of the memoryless communication channel.

Remark 5.8 In light of the previous remark where we do not need to consider the block error probability of fixed composition codes as an intermediate step, the ISP bound differs from the VF bound [109] (see Theorem 5.4) in the sense that the term $\frac{\log\binom{N+K-1}{K-1}}{N}$ is removed from $O_1(\frac{\ln N}{N}, x)$ (see (5.22)). Therefore, the shift in the rate of the error exponent of the ISP bound scales asymptotically like $\frac{1}{N}$ instead of $\frac{\ln N}{N}$ (see (5.21), (5.22) and (5.62)). Additionally, the derivation of the VF bound requires expurgation of the code to transform a lower bound on the maximal block error probability to a lower bound on the average block error probability. These differences

indicate a tightening of the pre-exponent of the ISP bound (as compared to the SP67 and VF bounds) which is expected to be especially pronounced for codes of small to moderate block lengths and also when the size of the channel input alphabet is large (as will be verified in Section 5.5).

Remark 5.9 The ISP bound is also applicable to symmetric channels with continuous output. When the ISP bound is applied to a memoryless symmetric channel with a continuous-output alphabet, the transition probability is replaced by a transition density function and the sums over the output alphabet are replaced by integrals. Note that these densities may include Dirac delta functions which appear at the points where the corresponding input distribution or the transition density function of the channel are discontinuous. Additionally, as explained in Appendix 5.A, the statement in Lemma 5.1 holds for general symmetric memoryless channels.

5.4 The 1959 Sphere-Packing Bound of Shannon and Improved Algorithms for Its Calculation

The 1959 sphere-packing (SP59) bound, derived by Shannon [89], provides a lower bound on the decoding error probability of an arbitrary block code whose transmission takes place over an AWGN channel. We begin this section by introducing the SP59 bound in its original form, along with asymptotic approximations in [89] which facilitate the estimation of the bound for large block lengths. We then review a theorem, introduced by Valembois and Fossorier [109], presenting a set of recursive equations which simplify the calculation of this bound. Both the original formula for the SP59 bound in [89] and the recursive method in [109] perform the calculations in the probability domain; this leads to various numerical difficulties of over and under flows when calculating the exact value of the bound for codes of block lengths of $N = 1000$ or more. In this section, we present an alternative approach which facilitates the calculation of the SP59 bound in the logarithmic domain. This eliminates the possibility of numerical problems in the calculation of the SP59 bound, regardless of the block length.

5.4.1 The 1959 Sphere-Packing Bound and Asymptotic Approximations

Consider a block code $\mathcal{C}$ of length N and rate R nats per channel use per dimension. It is assumed that all the codewords are mapped to signals with equal energy

(e.g., PSK modulation); hence, all the signals representing codewords lie on an N-dimensional sphere centered at the origin, but finer details of the modulation used are not taken into account in the derivation of the bound. This assumption implies that every Voronoi cell (i.e., the convex region containing all the points which are closer to the considered signal than to any other code signal) is a polyhedric cone which is limited by at most $\exp(NR) - 1$ hyper planes intersecting at the origin. As a measure of volume, Shannon introduced the solid angle of a cone which is defined to be the area of the sphere of unit radius cut out by the cone. Since the Voronoi cells partition the space $\mathbb{R}^N$, then the sum of their solid angles is equal to the area of an N-dimensional sphere of unit radius. The derivation of the SP59 bound relies on two main observations:

- Among the cones of a given solid angle, the lowest probability of error is obtained by the circular cone whose main axis passes through the origin and the signal point which represents the transmitted signal.
- In order to minimize the average decoding error probability, it is best to share the total solid angle equally among the $\exp(NR)$ Voronoi regions.

As a corollary of these two observations, it follows that the average block error probability cannot be smaller than the error probability which corresponds to the case where all the Voronoi regions are circular cones centered around the code signals with a common solid angle which is equal to a fraction of $\exp(-NR)$ of the solid angle of $\mathbb{R}^N$. The solid angle of a circular cone is given by the following lemma.

Lemma 5.2 [Solid Angle of a Circular Cone [89]] The solid angle of a circular cone of half angle θ in $\mathbb{R}^N$ is given by

$$\Omega_N(\theta) = \frac{2\pi^{\frac{N-1}{2}}}{\Gamma(\frac{N-1}{2})} \int_0^\theta (\sin\phi)^{N-2}\,\mathrm{d}\phi\,.$$

In particular, the solid angle of $\mathbb{R}^N$ is given by

$$\Omega_N(\pi) = \frac{2\pi^{\frac{N}{2}}}{\Gamma(\frac{N}{2})}\,.$$

Theorem 5.6 [The 1959 Sphere-Packing (SP59) Bound [89]] Assume that the transmission of an arbitrary block code of length N and rate R (in units of nats per channel use per dimension) takes place over an AWGN channel whose additive white Gaussian noise has a two-sided spectral density of $\frac{N_0}{2}$. Then, under ML decoding, the block error probability is lower bounded by

$$P_{\mathrm{e}}(\mathrm{ML}) > P_{\mathrm{SPB}}(N,\theta,A)\,, \quad A \triangleq \sqrt{\frac{2E_{\mathrm{s}}}{N_0}} \tag{5.65}$$

where E_{s} is the average energy per symbol, $\theta \in [0, \pi]$ satisfies the inequality $\exp(-NR) \leq \frac{\Omega_N(\theta)}{\Omega_N(\pi)}$,

$$P_{\mathrm{SPB}}(N, \theta, A) \triangleq \frac{(N-1)\exp\left(-\frac{NA^2}{2}\right)}{\sqrt{2\pi}} \int_{\theta}^{\frac{\pi}{2}} (\sin\phi)^{N-2} f_N(\sqrt{N}A\cos\phi)\,\mathrm{d}\phi + Q(\sqrt{N}A) \tag{5.66}$$

and

$$f_N(x) \triangleq \frac{1}{2^{\frac{N-1}{2}}\Gamma(\frac{N+1}{2})} \int_0^{\infty} z^{N-1} \exp\left(-\frac{z^2}{2} + zx\right) \mathrm{d}z\,, \quad \forall\, x \in \mathbb{R},\, N \in \mathbb{N}. \tag{5.67}$$

By assumption, the transmitted signal is represented by a point which lies on the N-dimensional sphere of radius $\sqrt{NE_s}$ and which is centered at the origin, and the Gaussian noise is additive. The value of $P_{\mathrm{SPB}}(N, \theta, A)$ on the RHS of (5.65) designates the probability that the received vector falls outside the N-dimensional circular cone of half angle θ whose main axis passes through the origin and the signal point which represents the transmitted signal. Hence, this function is monotonically decreasing in θ. The tightest lower bound on the decoding error probability, as given in (5.65), is therefore achieved for $\theta_1(N, R)$ which satisfies

$$\frac{\Omega_N\big(\theta_1(N,R)\big)}{\Omega_N(\pi)} = \exp(-NR). \tag{5.68}$$

In order to simplify the calculation of the SP59 bound, Shannon provided in [89] asymptotically tight upper and lower bounds on the ratio $\frac{\Omega_N(\theta)}{\Omega_N(\pi)}$.

Lemma 5.3 [Bounds on the Solid Angle [89]] The solid angle of a circular cone of half angle θ in the Euclidean space $\mathbb{R}^N$ satisfies the inequality

$$\frac{\Gamma(\frac{N}{2})(\sin\theta)^{N-1}}{2\Gamma(\frac{N+1}{2})\sqrt{\pi}\cos\theta}\left(1 - \frac{\tan^2\theta}{N}\right) \leq \frac{\Omega_N(\theta)}{\Omega_N(\pi)} \leq \frac{\Gamma(\frac{N}{2})(\sin\theta)^{N-1}}{2\Gamma(\frac{N+1}{2})\sqrt{\pi}\cos\theta}\,.$$

Corollary 5.2 [SP59 Bound (Cont.)] If θ^* satisfies the equality

$$\frac{\Gamma(\frac{N}{2})(\sin\theta^*)^{N-1}}{2\Gamma(\frac{N+1}{2})\sqrt{\pi}\cos\theta^*}\left(1 - \frac{\tan^2\theta^*}{N}\right) = \exp(-NR) \tag{5.69}$$

then $\frac{\Omega_N(\theta^*)}{\Omega_N(\pi)} \geq \exp(-NR)$, and therefore

$$P_{\mathrm{e}}(\mathrm{ML}) > P_{\mathrm{SPB}}(N, \theta^*, A). \tag{5.70}$$

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 $\frac{\Omega_N(\theta)}{\Omega_N(\pi)}$, this loss vanishes as $N \to \infty$. In [109], it was numerically observed that this loss is marginal even for relatively small values of NR; it was observed that this loss is smaller than 0.01 dB whenever the dimension of the code in bits is greater than 20, and it becomes smaller than 0.001 dB when the dimension exceeds 60 bits.

For large block lengths, the calculation of the SP59 bound becomes difficult in practice due to over and under flows in the floating-point operations. However, [89] presents some asymptotic formulas which give a good estimation of the bound for large enough block lengths. These approximations allow the calculation to be made in the logarithmic domain which eliminates the possibility of floating-point errors.

Theorem 5.7 [89]. Defining

$$\begin{aligned} G(\theta) &\triangleq \frac{A\cos\theta + \sqrt{A^2\cos^2\theta + 4}}{2} \\ E_L(\theta) &\triangleq \frac{A^2 - AG(\theta)\cos\theta - 2\ln\big(G(\theta)\sin\theta\big)}{2} \end{aligned}$$

then

$$P_{\mathrm{SPB}}(N,\theta,A) \geq \frac{\sqrt{N-1}}{6N(A+1)} \exp\left(\frac{3-(A+1)^2}{2}\right) \exp\big(-N\,E_L(\theta)\big). \tag{5.71}$$

This lower bound is valid for any block length N. However, the ratio of the left and right terms in (5.71) stays bounded away from one for all N.

A rather accurate approximation of $P_{\mathrm{SPB}}(N,\theta,A)$ was provided by Shannon in [89], but without a determined inequality. As a consequence, the following approximation is not a proven theoretical lower bound on the block error probability. For $N > 1000$, however, its numerical values become almost identical to those of the exact bound, thus giving a useful estimation for the lower bound.

Proposition 5.2 [89]. Using the notation of Theorem 5.7, if $\theta > \cot^{-1}(A)$, then

$$P_{\mathrm{SPB}}(N,\theta,A) \approx \frac{\alpha(\theta)\exp(-NE_L\big(\theta)\big)}{\sqrt{N}}$$

where

$$\alpha(\theta) \triangleq \left(\sqrt{\pi\,(1+G(\theta)^2)}\,\sin\theta\,\big(AG(\theta)\sin^2\theta - \cos\theta\big)\right)^{-1}.$$

5.4.2 A Recent Algorithm for Calculating the 1959 Sphere-Packing Bound

In [109, Section 2], Valembois and Fossorier review the SP59 bound and suggest a recursive algorithm to simplify its calculation. This algorithm is presented in the following theorem:

Theorem 5.8 [Recursive Equations for Simplifying the Calculation of the SP59 Bound] [109, Theorem 3]. The set of functions $\{f_N\}$ introduced in (5.67) can be expressed in the alternative form

$$f_N(x) = P_N(x) + Q_N(x)\exp(\frac{x^2}{2})\int_{-\infty}^{x}\exp(-\frac{t^2}{2})\,dt\,, \quad x \in \mathbb{R},\, N \in \mathbb{N} \tag{5.72}$$

where P_N and Q_N are two polynomials, determined by the same recursive equation for $N \geq 5$

$$\begin{aligned} P_N(x) &= \frac{2N-5+x^2}{N-1}\,P_{N-2}(x) - \frac{N-4}{N-1}\,P_{N-4}(x)\,, \\ Q_N(x) &= \frac{2N-5+x^2}{N-1}\,Q_{N-2}(x) - \frac{N-4}{N-1}\,Q_{N-4}(x) \end{aligned} \tag{5.73}$$

with the initial conditions

$$\begin{aligned} &P_1(x) = 0, \quad P_2(x) = \sqrt{\frac{2}{\pi}}, \quad P_3(x) = \frac{x}{2}, \quad P_4(x) = \sqrt{\frac{2}{\pi}}\frac{2+x^2}{3}, \\ &Q_1(x) = 1, \quad Q_2(x) = \sqrt{\frac{2}{\pi}}\,x, \quad Q_3(x) = \frac{1+x^2}{2}, \quad Q_4(x) = \sqrt{\frac{2}{\pi}}\frac{3x+x^3}{3}. \end{aligned}$$

By examining the recursive equations for P_N and Q_N in (5.73), it is observed that the coefficients of the higher powers of x vanish exponentially as N increases. When performing the calculation using double-precision floating-point numbers, these coefficients cause underflows when N is larger than several hundreds, and are replaced by zeros. Examining the expression for $P_{\mathrm{SPB}}(N,\theta,A)$ in (5.66), we observe that $f_N(x)$ (and therefore the polynomials $P_N(x)$ and $Q_N(x)$) are evaluated at $x \sim O(\sqrt{N})$. Hence, for large values of N, the replacement of the coefficients of the high powers of x by zeros causes a considerable inaccuracy in the calculation of P_{SPB} in (5.66).

Considering the integrand on the RHS of (5.66) reveals another difficulty in calculating the SP59 bound for large values of N. In this case, the term $f_N(\sqrt{N}A\cos\phi)$ becomes very large and causes overflows, while the value of the term $(\sin\phi)^{N-2}$ becomes very small and causes underflows; this creates a "$0\cdot\infty$" phenomenon when evaluating the integrand at the RHS of (5.66).

5.4.3 A Log-Domain Approach for Computing the 1959 Sphere-Packing Bound

In this section, we present a method which facilitates the entire calculation of the integrand on the RHS of (5.66) in the logarithmic domain, thus circumventing the numerical over and under flows which become problematic in the calculation of the SP59 bound for large block lengths. We begin our derivation by representing the set of functions $\{f_N\}$ defined in (5.67) as sums of exponents.

Proposition 5.3 The set of functions $\{f_N\}$ in (5.67) can be expressed in the form

$$f_N(x) = \sum_{j=0}^{N-1} \exp\big(d(N,j,x)\big)\,, \quad x \in \mathbb{R},\, N \in \mathbb{N}$$

where

$$\begin{aligned} d(N,j,x) \;\triangleq\; & \frac{x^2}{2} + \ln\Gamma\left(\frac{N}{2}\right) - \ln\Gamma\left(\frac{j}{2}+1\right) - \ln\Gamma(N-j) \\ & +(N-1-j)\ln\left(\sqrt{2}\,x\right) - \frac{\ln 2}{2} \\ & +\ln\left[1+(-1)^j\,\tilde{\gamma}\left(\frac{x^2}{2},\frac{j+1}{2}\right)\right], \qquad \begin{array}{c} N\in\mathbb{N},\, x\in\mathbb{R} \\ j=0,1\ldots,N-1 \end{array} \end{aligned} \tag{5.74}$$

and

$$\Gamma(a) \;\triangleq\; \int_0^\infty t^{a-1}\exp(-t)\,dt\,, \quad \mathrm{Re}(a)>0 \tag{5.75}$$

$$\tilde{\gamma}(x,a) \;\triangleq\; \frac{1}{\Gamma(a)}\int_0^x t^{a-1}\exp(-t)\,dt\,, \quad x\in\mathbb{R},\, \mathrm{Re}(a)>0 \tag{5.76}$$

designate the complete and incomplete Gamma functions, respectively.

Proof: The proof is given in Appendix 5.C. ■

Remark 5.10 It is noted that the exponents $d(N,j,x)$ in (5.74) are readily calculated by using standard mathematical functions. The function which calculates the natural logarithm of the Gamma function is implemented in the MATLAB software by `gammaln`, and in the Mathematica software by `LogGamma`. The incomplete Gamma function $\tilde{\gamma}$ is implemented in MATLAB by `gammainc` and in Mathematica by `GammaRegularized`.

In order to perform the entire calculation of the function f_N in the logarithmic domain, we employ the function

$$\max{}^*(x_1,\ldots,x_m) \triangleq \ln\left(\sum_{i=1}^m \exp(x_i)\right)\,, \quad m\in\mathbb{N}, \quad x_1,\ldots,x_m\in\mathbb{R} \tag{5.77}$$

which is commonly used for the implementation of the log-domain BCJR algorithm. The function $\max^*$ can be calculated in the logarithmic domain using the recursive equation

$$\max{}^*(x_1,\ldots,x_{m+1}) = \max{}^*\big(\max{}^*(x_1,\ldots,x_m),x_{m+1}\big)\,, \qquad \begin{array}{l} m\in\mathbb{N}\setminus\{1\},\\ x_1,\ldots,x_{m+1}\in\mathbb{R}\end{array}$$

with the initial condition

$$\max{}^*(x_1,x_2) = \max(x_1,x_2) + \ln\Big(1+\exp\big(-|x_1-x_2|\big)\Big)\,.$$

Combining Proposition 5.3 and the definition of the function $\max^*$ in (5.77), we get a method for calculating the set of functions $\{f_N\}$ in the logarithmic domain.

Corollary 5.3 The set of functions $\{f_N\}$ defined in (5.67) can be rewritten in the form

$$f_N(x) = \exp\Big[\max{}^*\big(d(N,0,x),d(N,1,x),\ldots,d(N,N-1,x)\big)\Big] \tag{5.78}$$

where $d(N,j,x)$ is introduced in (5.74).

By combining (5.66) and (5.78), one gets the following theorem which provides an efficient algorithm for the calculation of the SP59 bound in the log domain.

Theorem 5.9 [Logarithmic domain calculation of the SP59 bound] The term $P_{\rm SPB}(N,\theta,A)$ on the RHS of (5.70) can be rewritten as

$$\begin{aligned} P_{\rm SPB}(N,\theta,A) \;=\; & \int_\theta^{\frac{\pi}{2}} \exp\Bigg[\ln(N-1) - \frac{NA^2}{2} - \frac{1}{2}\ln(2\pi) + (N-2)\ln\sin\phi \\ & \qquad + \max{}^*\Big(d(N,0,\sqrt{N}A\cos\phi),\ldots,d(N,N-1,\sqrt{N}A\cos\phi)\Big)\Bigg]d\phi \\ & + Q(\sqrt{N}A)\,, \qquad N\in\mathbb{N},\ \theta\in[0,\frac{\pi}{2}],\ A\in\mathbb{R}^+ \end{aligned}$$

where the function d is introduced in (5.74).

Using Theorem 5.9, it is easy to calculate the exact value of the SP59 lower bound for very large block lengths.

5.5 Numerical Results for Sphere-Packing Bounds

This section presents some numerical results which serve to exemplify the improved tightness of the ISP bound derived in Section 5.3. We consider performance bounds for coherent detection of M-ary PSK block coded modulation where the signals are

transmitted over fully interleaved fading channels, and it is assumed that perfect side information of the fading samples is available at the receiver. As special cases, the fully interleaved Rayleigh-fading channel and the AWGN channel are considered. For M-ary PSK modulated signals whose transmission takes place over the AWGN channel, the ISP bound is compared to the SP59 bound (which is revisited in Section 5.4) and to some upper bounds on the decoding error probability. As a representative of the class of discrete memoryless and symmetric channels, the binary erasure channel (BEC) is considered. All the bounds are compared in this section to computer simulations for the performance of modern error-correcting codes using practical decoding algorithms.

5.5.1 Performance Bounds for M-ary PSK Block Coded Modulation over Fully Interleaved Fading Channels

The ISP bound in Section 5.3 is particularized here to M-ary PSK block coded modulation schemes whose transmission takes place over fully interleaved fading channels, where it is assumed that the received signals are coherently detected and the fading samples are perfectly known at the receiver. For simplicity of notation, we treat the channel inputs and outputs as two-dimensional real vectors, and not as complex numbers. Let $M = 2^p$ (where $p \in \mathbb{N}$) be the size of the constellation for the PSK modulation, and denote the input to the channel by $\mathbf{X} = (X_1, X_2)$ where the possible input values are given by

$$\mathbf{x}_k = (\cos\theta_k, \sin\theta_k), \quad \theta_k \triangleq \frac{(2k+1)\pi}{M}, \quad k = 0, 1, \ldots, M-1. \tag{5.79}$$

We denote the channel output by $(\mathbf{Y}, A) = (Y_1, Y_2, A)$ where A is a fading sample which is distributed according to some distribution (or density function) p_A, $\mathbf{Y} = A\mathbf{X} + \mathbf{N}$, and $\mathbf{N} = (N_1, N_2)$ is an additive Gaussian random vector with i.i.d. components with zero-mean and variance σ^2. The channel input, fading sample and additive noise are statistically independent. The conditional *pdf* of the channel output, given the transmitted symbol $\mathbf{X}_k$, is given by

$$p_{\mathbf{Y},A|\mathbf{X}}(\mathbf{y}, a|\mathbf{x}_k) = \frac{p_A(a)}{2\pi\sigma^2} \exp\left(-\frac{\|\mathbf{y} - a\mathbf{x}_k\|^2}{2\sigma^2}\right), \quad \mathbf{y} \in \mathbb{R}^2, a \in \mathbb{R}^+ \tag{5.80}$$

where $\|\cdot\|$ designates the L_2 norm. Due to the symmetry of the additive noise and the fact that the fading samples are fully known at the receiver, the phase of the fading coefficient can be eliminated at the receiver; hence, the fading is treated as a non-negative (real) random variable. Due to the channel interleaver, the fading

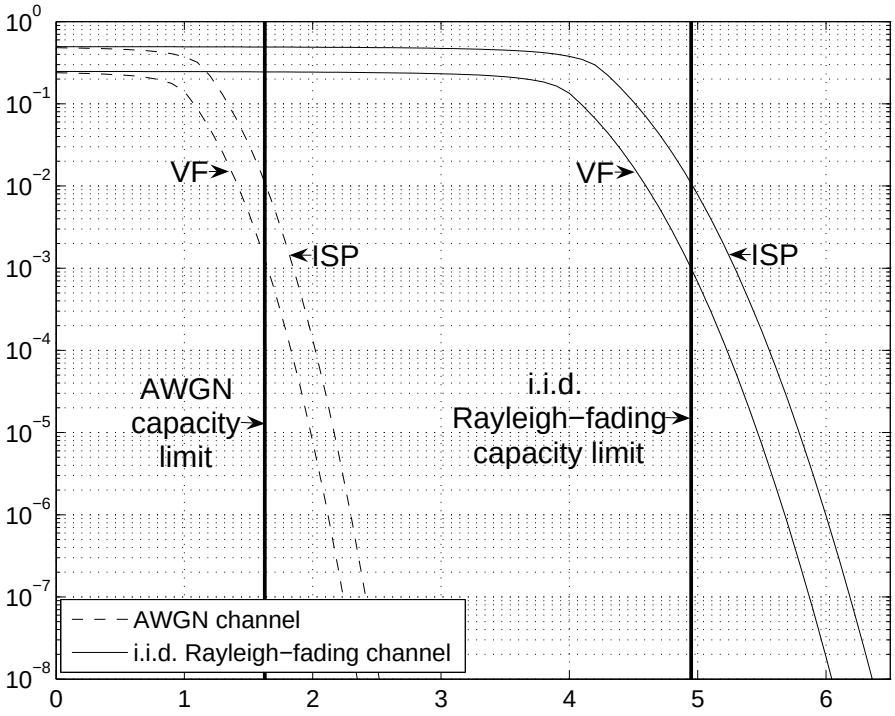

Figure 5.1: A comparison between lower bounds on the ML decoding error probability for block codes of length $N = 1024$ bits and code rate of $0.75 \frac{\text{bits}}{\text{channel use}}$. This figure refers to BPSK modulated signals whose transmission takes place over fully-interleaved (i.i.d.) Rayleigh-fading and AWGN channels. We compare the Valembois-Fossorier (VF) [109] bound and the improved sphere-packing (ISP) bound derived in Section 5.3.

coefficients are i.i.d. random variables so the channel is indeed memoryless. Closed-form expressions for the function μ_0 and its first two derivatives w.r.t. s (while holding f_s fixed) are derived in Appendix 5.B.1 and are used for the calculation of both the VF and ISP bounds. Further details on BPSK modulated signals transmitted over fully interleaved fading channels, including expressions for the capacity, cutoff rate and various bounds on the decoding error probability, are provided in [78] and references therein.

Figure 5.1 compares the VF bound [109] and the ISP bound derived in Section 5.3. The comparison refers to block codes of length 1024 bits and rate $0.75 \frac{\text{bits}}{\text{channel use}}$ which employ BPSK modulation. Two communication channels are considered: The AWGN channel, which can be viewed as a fading channel where the fading samples are set to 1 (i.e., $p_A(a) = \delta(a-1)$ where δ designates the Dirac delta function), and the fully interleaved Rayleigh-fading channel, where

$$p_A(a) = 2a \exp(-a^2), \quad a \geq 0.$$

The plot also depicts the capacity limit bound (CLB) for these two channels (calculated from [78, Eq. (2)]).[1] It is observed that the ISP bound outperforms the VF bound for both channels and that the gap is wider for the Rayleigh-fading channel. For a block error probability of 10^{-5}, the ISP bound provides gains of about 0.19 and 0.33 dB over the VF bound and gaps of 0.54 dB and 0.84 dB to the channel capacity for the AWGN and Rayleigh-fading channels, respectively. Also, for both channels the ISP bound is more informative than the CLB for block error probabilities below 10^{-2} while the VF bound requires block error probabilities below 10^{-3} to outperform the capacity limit.

5.5.2 Performance Bounds for M-ary PSK Block Coded Modulation over the AWGN Channel

The ISP bound is particularized in Section 5.5.1 to M-ary PSK block coded modulation schemes whose transmission takes place over fully interleaved fading channels, where the received signals are coherently detected and the fading samples are fully known at the receiver. A special case of this model is the AWGN channel. The closed-form expressions for the function μ_0 and its first two derivatives w.r.t. s (while holding f_s fixed) are given in Appendix 5.B.2. The SP59 bound [89] provides a lower bound on the decoding error probability for the considered case, since the modulated signals have equal energy and are transmitted over the AWGN channel. In the following, we exemplify the use of these lower bounds. They are also compared to the RCB of Gallager [31], and the tangential-sphere upper bound (TSB) of Poltyrev [70] when applied to random block codes. This serves for the study of the tightness of the ISP bound, as compared to other upper and lower bounds. The numerical results shown in this section indicate that the recent variants of the SP67 bound provide an interesting alternative to the SP59 bound which is commonly used in the literature as a measure for the sub-optimality of codes transmitted over the AWGN channel (see, e.g., [27, 44, 55, 79, 98, 109, 115]). Moreover, the advantage of the ISP bound over the VF bound in [109] is exemplified in this section.

Figure 5.2 compares the SP59 bound [89], the VF bound [109], and the ISP bound derived in Section 5.3. The comparison refers to block codes of length 500 bits and rate $0.8 \frac{\text{bits}}{\text{channel use}}$ which are BPSK modulated and transmitted over an AWGN channel. The plot also depicts the RCB of Gallager [31], the TSB ([36, 70]), and the capacity limit bound (CLB). It is observed from this figure that even for relatively short block

[1] Although the CLB refers to the asymptotic case where the block length tends to infinity, it is plotted in [109] and here as a reference, in order to examine whether the improvement in the tightness of the ISP is for rates above or below capacity.

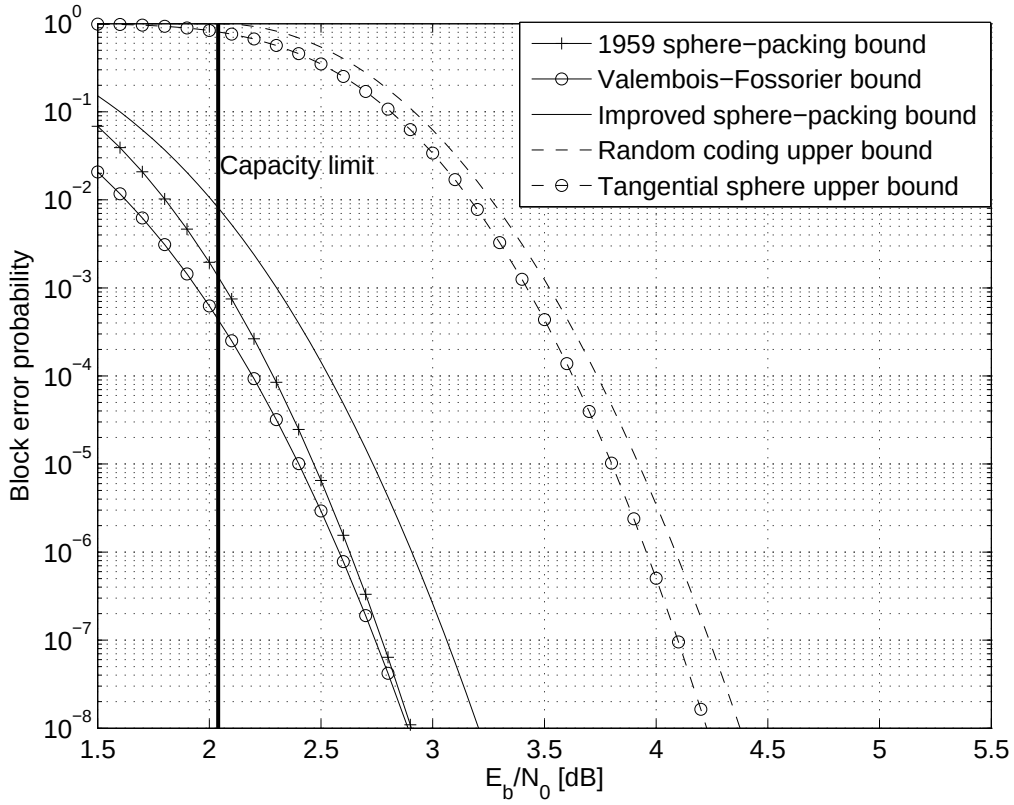

Figure 5.2: A comparison between upper and lower bounds on the ML decoding error probability for block codes of length $N = 500$ bits and code rate of $0.8 \frac{\text{bits}}{\text{channel use}}$. This figure refers to BPSK modulated signals whose transmission takes place over an AWGN channel. The compared bounds are the 1959 sphere-packing (SP59) bound of Shannon [89], the Valembois-Fossorier (VF) bound [109], the improved sphere-packing (ISP) bound derived in Section 5.3, the random-coding upper bound (RCB) of Gallager [31], and the tangential-sphere bound (TSB) [36, 70] when applied to fully random block codes with the above block length and rate.

lengths, the ISP bound outperforms the SP59 bound for block error probabilities below 10^{-1} (this issue will be discussed later in this section). For a block error probability of 10^{-5}, the ISP bound provides gains of about 0.26 and 0.33 dB over the SP59 and VF bounds, respectively. For these code parameters, the TSB provides a tighter upper bound on the block error probability of random codes, as compared to the RCB of Gallager; e.g., the gain of the TSB over the Gallager bound is about 0.2 dB for a block error probability of 10^{-5}. Note that the Gallager bound is tighter than the TSB for fully random block codes of large enough block lengths, as the latter bound does not reproduce the random-coding error exponent for the AWGN channel [70]. However, Figure 5.2 exemplifies the advantage of the TSB over the Gallager bound,

when applied to random block codes of relatively short block lengths; this advantage is especially pronounced for low code rates where the gap between the error exponents of these two bounds is marginal (see [79, p. 67]), but it is also reflected from Figure 5.2 for BPSK modulation with a code rate of $0.8\frac{\text{bits}}{\text{channel use}}$. The gap between the TSB and the ISP bound, as upper and lower bounds respectively, is less than 1.2 dB for all block error probabilities lower than 10^{-1}. Also, the ISP bound is more informative than the CLB for block error probabilities below $8 \cdot 10^{-3}$ while the SP59 and VF bounds require block error probabilities below $1.5 \cdot 10^{-3}$ and $5 \cdot 10^{-4}$, respectively, to outperform the capacity limit.

Figure 5.3 presents a comparison of the SP59, VF and ISP bounds referring to short block codes which are QPSK modulated and transmitted over the AWGN channel. The plots also depict the RCB, the TSB and CLB; in these plots, the ISP bound outperforms the SP59 bound for all block error probabilities below $4 \cdot 10^{-1}$ (this result is consistent with the upper plot of Figure 5.7). In the upper plot of Figure 5.3, which corresponds to a block length of 1024 bits (i.e., 512 QPSK symbols) and a rate of $1.5\frac{\text{bits}}{\text{channel use}}$, it is shown that the ISP bound provides gains of about 0.25 and 0.37 dB over the SP59 and VF bounds, respectively, for a block error probability of 10^{-5}. The gap between the ISP lower bound and the RCB is 0.78 dB for all block error probabilities lower than 10^{-1}. In the lower plot of Figure 5.3 which corresponds to a block length of 300 bits and a rate of $1.8\frac{\text{bits}}{\text{channel use}}$, the ISP bound significantly improves the SP59 and VF bounds; for a block error probability of 10^{-5}, the improvement in the tightness of the ISP over the SP59 and VF bounds is 0.8 and 1.13 dB, respectively. Additionally, the ISP bound is more informative than the CLB for block error probabilities below $3 \cdot 10^{-3}$, where the SP59 and VF bound outperform the CLB only for block error probabilities below $3 \cdot 10^{-6}$ and $5 \cdot 10^{-8}$, respectively. For fully random block codes of length $N = 300$ and rate $1.8\frac{\text{bits}}{\text{channel use}}$ which are QPSK modulated with Gray's mapping and transmitted over the AWGN channel, the TSB is tighter than the RCB (see the lower plot in Figure 5.3 and the explanation referring to Figure 5.2). The gap between the ISP bound and the TSB in this plot is about 1.5 dB for a block error probability of 10^{-5} (as compared to gaps of 2.3 dB (2.63 dB) between the TSB and the SP59 (VF) bound).

Figure 5.4 presents a comparison of the bounds for codes of block length 5580 bits and 4092 information bits, where both QPSK (upper plot) and 8-PSK (lower plot) constellations are considered. The modulated signals correspond to 2790 and 1680 symbols, respectively, so the code rates for these constellations are 1.467 and 2.2 bits per channel use, respectively. For both constellations, the two considered SP67-based

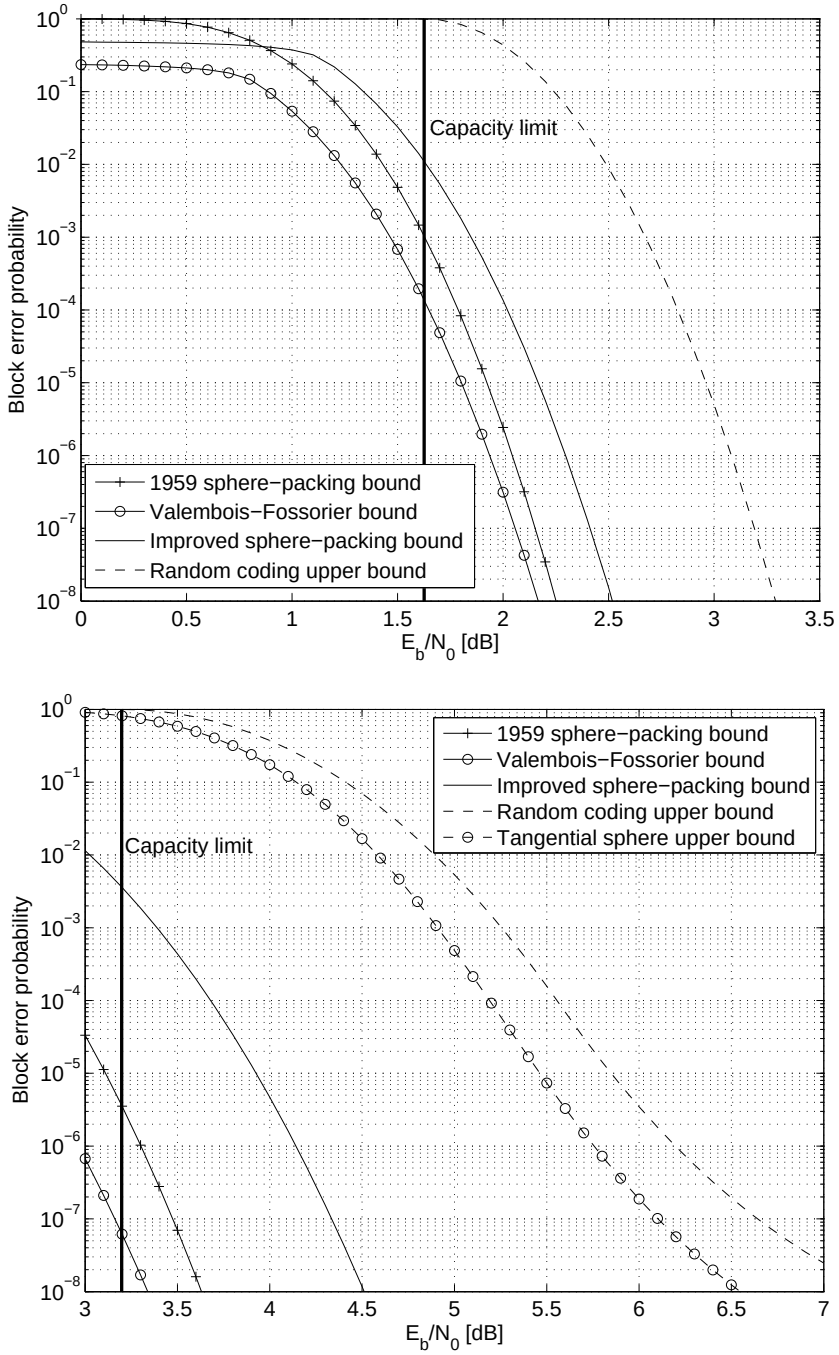

Figure 5.3: A comparison between upper and lower bounds on the ML decoding error probability, referring to short block codes which are QPSK modulated and transmitted over the AWGN channel. The compared lower bounds are the 1959 sphere-packing (SP59) bound of Shannon [89], the Valembois-Fossorier (VF) bound [109], and the improved sphere-packing (ISP) bound; the compared upper bounds are the random-coding upper bound (RCB) of Gallager [31] and the tangential-sphere bound (TSB) of Poltyrev [70]. The upper plot refers to block codes of length $N = 1024$ which are encoded by 768 information bits (so the rate is $1.5 \frac{\text{bits}}{\text{channel use}}$), and the lower plot refers to block codes of length $N = 300$ which are encoded by 270 bits whose rate is therefore $1.8 \frac{\text{bits}}{\text{channel use}}$.

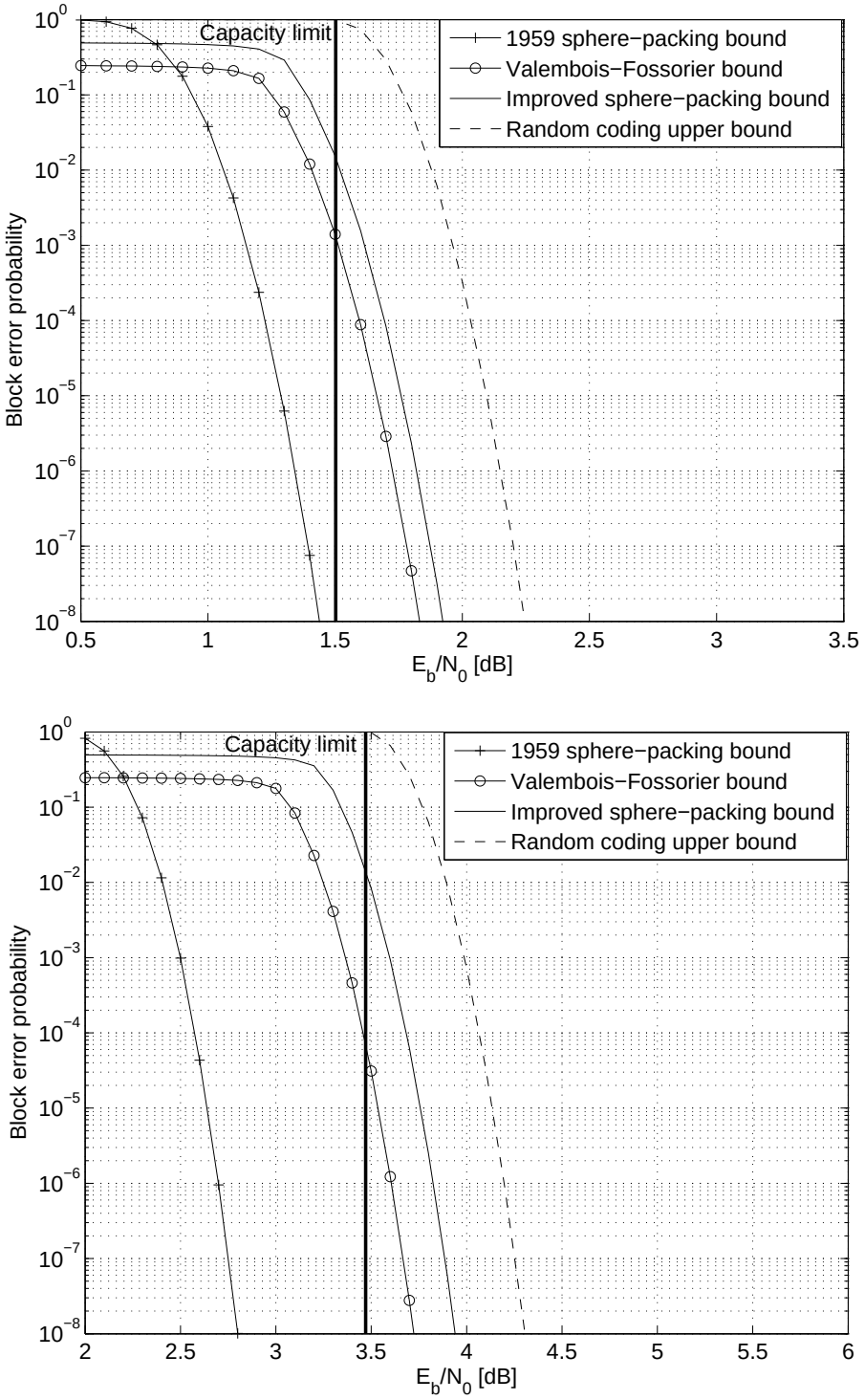

Figure 5.4: A comparison of upper and lower bounds on the ML decoding error probability for block codes of length $N = 5580$ bits and information block length of 4092 bits. This figure refers to QPSK (upper plot) and 8-PSK (lower plot) modulated signals whose transmission takes place over an AWGN channel; the rates in this case are 1.467 and $2.200 \frac{\text{bits}}{\text{channel use}}$, respectively. The compared bounds are the 1959 sphere-packing (SP59) bound of Shannon [89], the Valembois-Fossorier (VF) bound [109], the improved sphere-packing (ISP) bound, and the random-coding upper bound (RCB) of Gallager [31].

bounds (i.e., the VF and ISP bounds) outperform the SP59 for all block error probabilities below $2 \cdot 10^{-1}$; the ISP bound provides gains of 0.1 and 0.22 dB over the VF bound for the QPSK and 8-PSK constellations, respectively. For both modulations, the gap between the ISP lower bound and the RCB of Gallager does not exceed 0.4 dB. In [23], Divsalar and Dolinar design codes with the considered parameters by using concatenated Hamming and accumulate codes. They also present computer simulations of the performance of these codes under iterative decoding, when the transmission takes place over the AWGN channel and several common modulation schemes are applied. For a block error probability of 10^{-4}, the gap between the simulated performance of these codes under iterative decoding, and the ISP lower bound, which gives an ultimate lower bound on the block error probability of optimally designed codes under ML decoding, is approximately 1.4 dB for QPSK and 1.6 dB for 8-PSK signaling. This provides an indication on the performance of codes defined on graphs and their iterative decoding algorithms, especially in light of the feasible complexity of the decoding algorithm which is linear in the block length. To conclude, it is reflected from the results plotted in Figure 5.4 that a gap of about 1.5 dB between the ISP lower bound and the performance of the iteratively decoded codes in [23] is mainly due to the imperfectness of these codes and their sub-optimal iterative decoding algorithm; this conclusion follows in light of the fact that for random codes of the same block length and rate, the gap between the ISP bound and the RCB is reduced to less than 0.4 dB.

While it was shown in Section 5.3 that the ISP bound is uniformly tighter than the VF bound (which in turn is uniformly tighter than the SP67 bound [87]), no such relations are shown between the SP59 bound and the recent improvements on the SP67 bound (i.e., the VF and ISP bounds). Figure 5.5 presents regions of code rates and block lengths for which the ISP bound outperforms the SP59 bound and the CLB; it refers to BPSK modulated signals transmitted over the AWGN channel and considers block error probabilities of 10^{-4}, 10^{-5} and 10^{-6}. It is reflected from this figure that for any rate $0 < R < 1$, there exists a block length $N = N(R)$ such that the ISP bound outperforms the SP59 bound for block lengths larger than $N(R)$; the same property also holds for the VF bound, but the value of $N(R)$ depends on the considered SP67-based bound, and it becomes significantly larger in the comparison of the VF and SP59 bounds. It is also observed that the value $N(R)$ is monotonically decreasing with R, and it approaches infinity as we let R tend to zero. An intuitive explanation for this behavior can be given by considering the capacity limits of the binary-input and the energy-constrained AWGN channels. For any value $0 \leq C < 1$, denote by $\frac{E_{\mathrm{b},1}(C)}{N_0}$ and $\frac{E_{\mathrm{b},2}(C)}{N_0}$ the values of $\frac{E_{\mathrm{b}}}{N_0}$ required to achieve a channel capacity of

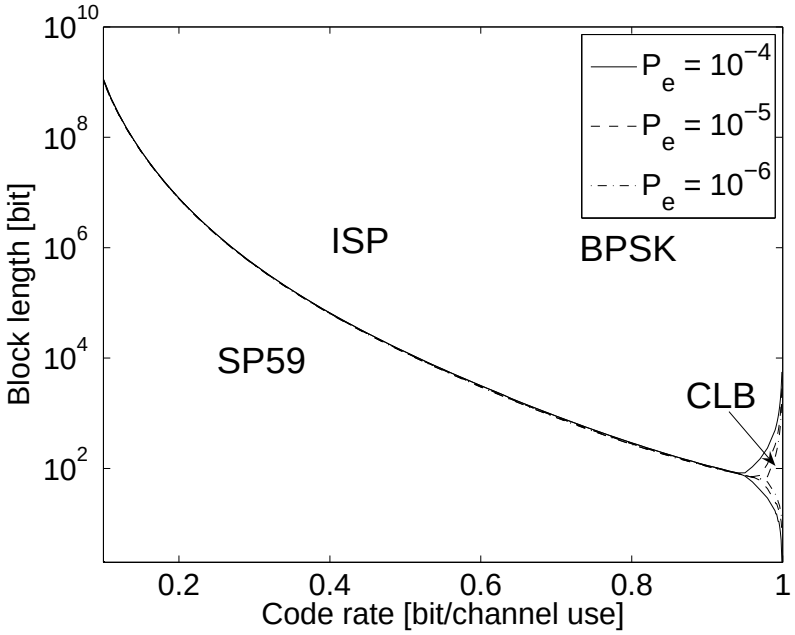

Figure 5.5: Regions in the two-dimensional space of code rate and block length, where a bound is better than the two others for three different targets of block error probability (P_e). The figure compares the tightness of the 1959 sphere-packing (SP59) bound of Shannon [89], the improved sphere-packing (ISP) bound, and the capacity-limit bound (CLB). The plot refers to BPSK modulated signals whose transmission takes place over the AWGN channel, and the considered code rates lie in the range between 0.1 and $1 \frac{\text{bits}}{\text{channel use}}$.

C bits per channel use for the binary-input and the energy-constraint AWGN channels, respectively (note that in the latter case, the input distribution which achieves capacity is also Gaussian). For any $0 \leq C < 1$, clearly $\frac{E_{\mathrm{b},1}(C)}{N_0} \geq \frac{E_{\mathrm{b},2}(C)}{N_0}$; however, the difference between these values is monotonically increasing with the capacity C, and, on the other hand, this difference approaches zero as we let C tend to zero. Since the SP59 bound only constrains the signals to be of equal energy, it gives a measure of performance for the energy-constrained AWGN channel, where the SP67-based bounds consider the actual modulation and therefore refer to the binary-input AWGN channel. As the code rates become higher, the difference in the ultimate performance between the two channels is larger, and therefore the SP67-based bounding techniques outperform the SP59 bound for smaller block lengths. For low code rates, the difference between the channels is reduced, and the SP59 outperforms the SP67-based bounding techniques even for larger block lengths due to the superior bounding technique which is specifically tailored for the AWGN channel.

Figure 5.6 presents regions of code rates and block lengths for which the VF

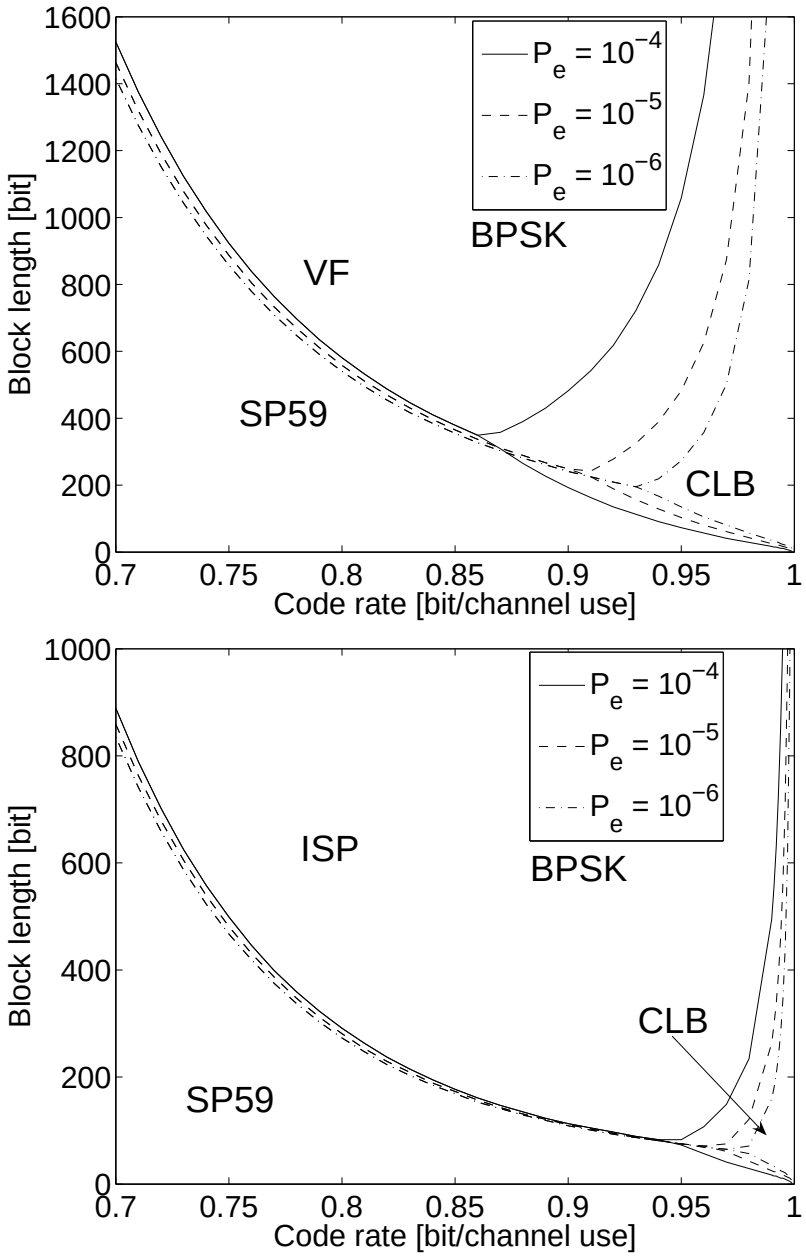

Figure 5.6: Regions in the two-dimensional space of code rate and block length, where a bound is better than the two others for three different targets of block error probability (P_e). The figure compares the tightness of the 1959 sphere-packing (SP59) bound of Shannon [89], the capacity-limit bound (CLB), and the Valembois-Fossorier (VF) bound [109] (upper plot) or the improved sphere-packing (ISP) bound in Section 5.3 (lower plot). The plots refer to BPSK modulated signals whose transmission takes place over the AWGN channel, and the considered code rates lie in the range between 0.70 and $1 \frac{\text{bits}}{\text{channel use}}$.

bound (upper plot) and the ISP bound (lower plot) outperform the CLB and the SP59 bound when the signals are BPSK modulated and transmitted over the AWGN channel; block error probabilities of 10^{-4}, 10^{-5} and 10^{-6} are examined. This figure focuses on high code rates, where the performance of the SP67-based bounds and their advantage over the SP59 bound is most appealing. From Figure 5.6, we have that for a code rate of 0.75 bits per channel use and a block error probability of 10^{-6}, the VF bound is tighter than the SP59 for block lengths exceeding 850 bits while the ISP bound reduces this value to 450 bits; moreover, when increasing the rate to 0.8 bits per channel use, the respective minimal block lengths reduce to 550 and 280 bits for the VF and ISP bounds, respectively. Figure 5.7 shows regions of code rates and block lengths where the ISP outperforms the CLB and SP59 bounds for QPSK (upper plot) and 8-PSK (lower plot) modulations. Comparing the lower plot of Figure 5.6 which refers to BPSK modulation with the upper plot of Figure 5.7 which refers to QPSK modulation, one can see that the two graphs are identical (when accounting for the doubling of the rate due to the use of both real and imaginary dimensions in the QPSK modulation). This is due to the fact that QPSK modulation poses no additional constraints on the channel and in fact, the real and imaginary planes can be serialized and decoded as in BPSK modulation. However, this property does not hold when replacing the ISP bound by the VF bound; this is due to the fact that the VF bound considers a fixed composition subcode of the original code and the increased size of the alphabet causes a greater loss in the rate for QPSK modulation. When comparing the two plots of Figure 5.7, it is evident that the minimal value of the block length for which the ISP bound becomes better than the SP59 bound decreases as the size of the input alphabet is increased (when the rate is measured in information bits per code bit). An intuitive justification for this phenomenon is attributed to the fact that referring to the constellation points of the M-ary PSK modulation, the mutual information between the code symbols in each dimension of the QPSK modulation is zero, while as the spectral efficiency of the PSK modulation is increased, the mutual information between the real and imaginary parts of each signal point is increased; thus, as the spectral efficiency is increased, this poses a stronger constraint on the possible positioning of the equal-energy signal points on the N-dimensional sphere. This intuition suggests an explanation for the reason why as the spectral efficiency is increased, the advantage of the ISP bound over the SP59 bound (which does not take into account the modulation scheme) holds even for smaller block lengths. This effect is expected to be more subtle for the VF bound since a larger size of the input alphabet decreases the rate for which the error exponent is evaluated (see (5.22)).

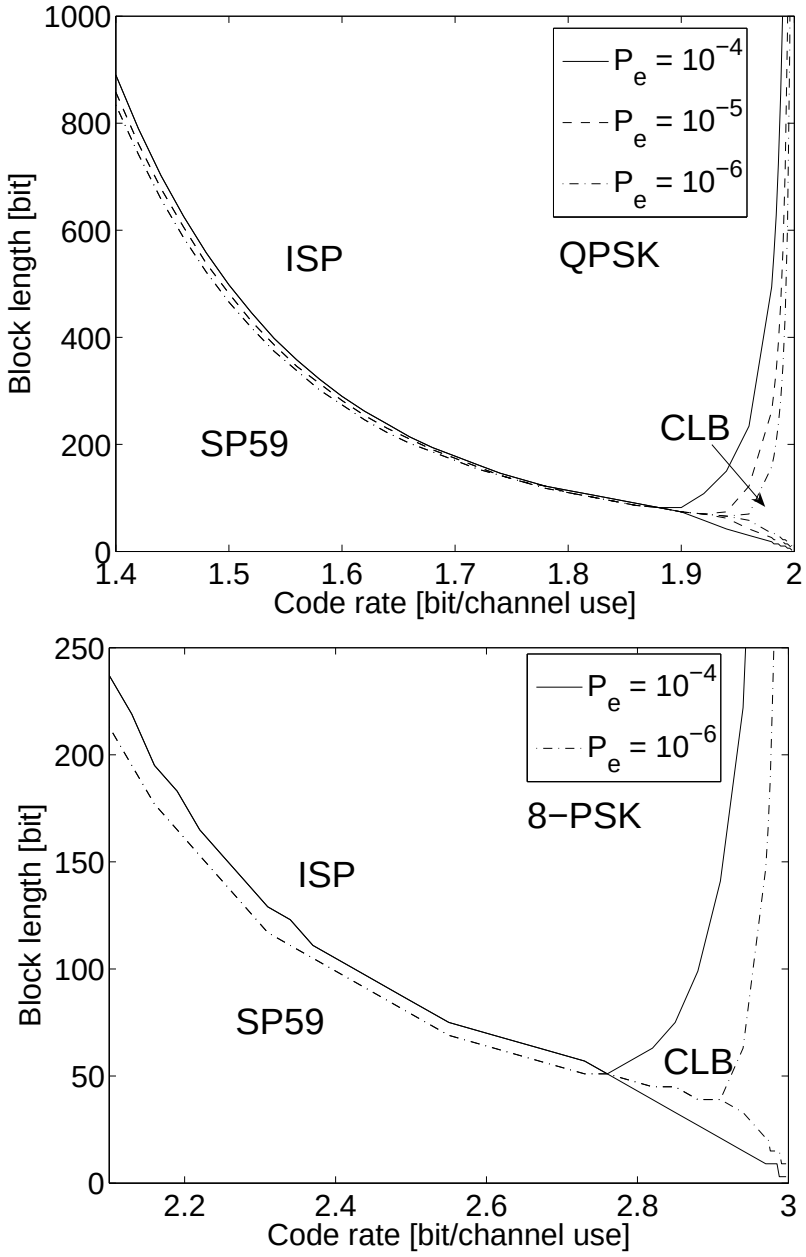

Figure 5.7: Regions in the two-dimensional space of code rate and block length, where a bound is better than the two others for different targets of block error probability (P_e). The figure compares the tightness of the 1959 sphere-packing (SP59) bound of Shannon [89], the improved sphere-packing (ISP) bound, and the capacity-limit bound (CLB). The plots refer to QPSK (upper plot) and 8-PSK (lower plot) modulated signals whose transmission takes place over the AWGN channel; the considered code rates lie in the range between 1.4 and $2\frac{\text{bits}}{\text{channel use}}$ for the QPSK modulated signals and between 2.1 and $3\frac{\text{bits}}{\text{channel use}}$ for the 8-PSK modulated signals.

5.5.3 Performance Bounds for the Binary Erasure Channel

In recent years, several families of code ensembles defined on graphs have been constructed and demonstrated to achieve the capacity of the BEC under iterative decoding with low complexity (see, e.g., [51], [64] and [93]). These low-complexity and capacity-achieving ensembles for the BEC motivate a study of the performance of iteratively decoded codes defined on graphs for moderate block lengths (see, e.g.,

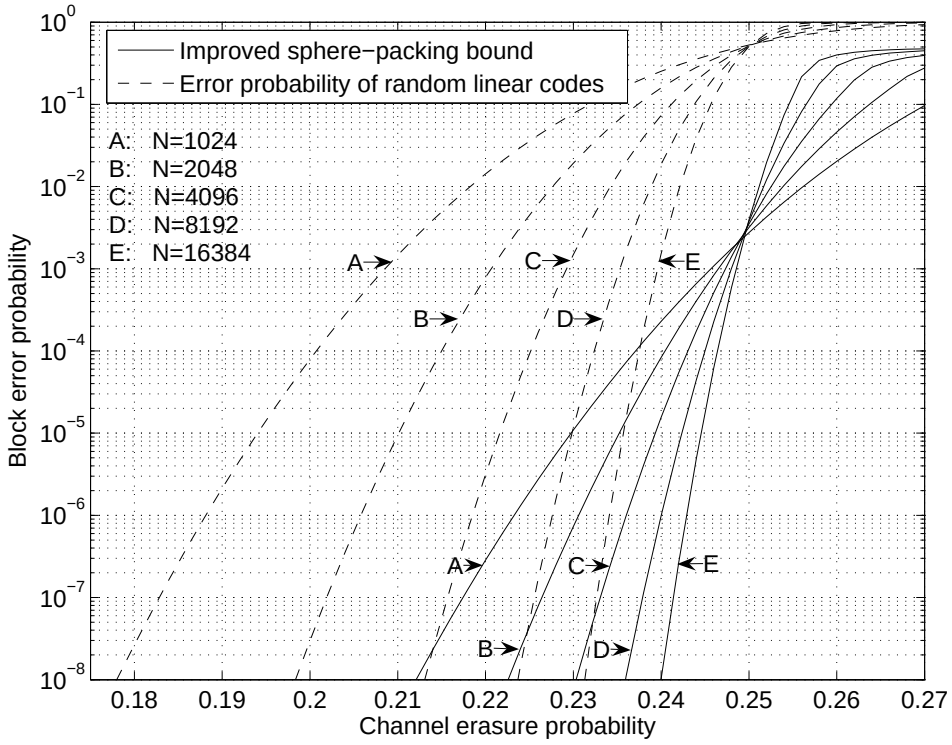

Figure 5.8: A comparison of the improved sphere-packing (ISP) lower bound from Section 5.3 and the exact decoding error probability of random binary linear block codes under ML decoding where the transmission takes place over the BEC (see [22, Eq. (3.2)]). The code rate examined is $0.75 \frac{\text{bits}}{\text{channel use}}$ and the block lengths are $N = 1024, 2048, 4096, 8192$ and 16384 bits.

[106]). In Figure 5.8, we compare the ISP lower bound and the exact block error probability of random linear block codes transmitted over the BEC as given in [22, Eq. (3.2)]. The figure refers to codes of rate 0.75 bits per channel use and various block lengths. It can be observed that for a block length of 1024 bits, the difference in the channel erasure probability for which the RCB and the ISP bound achieve a block error probability of 10^{-5} is 0.035 while for a block length of 16384 bits, this gap is decreased to 0.009. This yields that the ISP bound is reasonably tight, and also suggests that this bound can be used in order to assess the imperfectness of turbo-like codes even for moderate block lengths.

5.5.4 Minimal Block Length as a Function of Performance

In a wide range of applications, the system designer needs to design a communication system which fulfills several requirements on the available bandwidth, acceptable delay for transmitting and processing the data while maintaining a certain fidelity criterion in reconstructing the data (e.g., the block error probability needs to be below a certain threshold). In this setting, one wishes to design a code which satisfies the delay constraint (i.e., the block length is limited) while adhering to the required performance over the given channel. By fixing the communication channel model, code rate (which is related to the bandwidth expansion caused by the error-correcting code) and the block error probability, sphere-packing bounds are transformed into lower bounds on the minimal block length required to achieve the desired block error probability at a certain gap to capacity using an arbitrary block code and decoding algorithm. Similarly, by fixing these parameters, upper bounds on the error probability of random codes under ML decoding are transformed into upper bounds on the block length required for ML decoded random codes to achieve a desired block error probability on a given communication channel.

In this section, we consider some practically decodable codes taken from some recent papers ([5], [25], [26], [97], [99], [108]). We examine the gap between channel capacity and the $\frac{E_{\mathrm{b}}}{N_0}$ for which they achieve a required block error probability as a function of the block length of these codes. The performance of these specific codes under their practical decoding algorithms is compared to the sphere-packing bounds and also to upper bounds on the error probability of random block codes; these bounds serve here as lower and upper bounds, respectively, on the block length required to achieve a given block error probability and code rate on a given channel using an optimal block code and decoding algorithm. Comparing the performance of specific codes and decoding algorithms to the information-theoretic limitations provided by the sphere-packing bounds, enables one to deduce how far in terms of delay is a practical system from the fundamental limitations of information theory.

Figure 5.9 considers some block codes of rate $\frac{1}{2}$ bits per channel use which are BPSK modulated and transmitted over the AWGN channel. The plot depicts the gap to capacity in dB for which these codes achieve block error probabilities of 10^{-4} and 10^{-5} under their practical decoding algorithms as a function of their block length. As a reference, this figure also plots lower bounds on the block length which stem from the SP59 and ISP bounds, and upper bounds on the block length of fully random binary block codes which are based on the RCB of Gallager [31] and the TSB of Poltyrev [70]; these bounds refer to a block error probability of 10^{-5}. For large enough block lengths, the RCB provides a tighter upper bound on the achievable gap to capacity

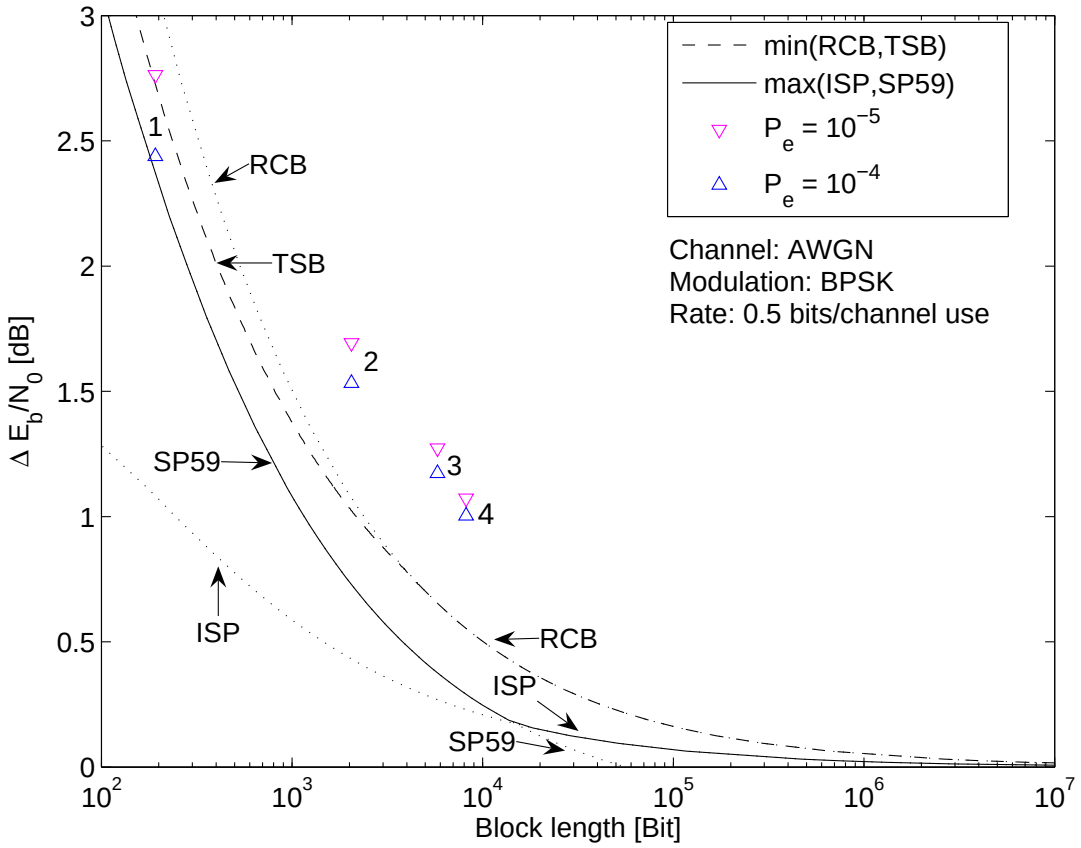

Figure 5.9: This figure refers to the tradeoff between the block length and the gap to capacity of error-correcting codes which are BPSK modulated and transmitted over an AWGN channel. The horizontal axis refers to the block length of the codes, and the vertical axis refers to the gap, measured in decibels, between the channel capacity and the energy per bit to spectral noise density ($\frac{E_b}{N_0}$) which is required to obtain a block error probability P_e (set either to 10^{-4} or 10^{-5}). The considered rate of all the codes is one-half bit per channel use. The minimal gap to capacity which is required for achieving a block error probability of 10^{-5} is depicted via bounds: the upper bound is calculated using the random-coding bound (RCB) of Gallager [31] and the tangential-sphere bound (TSB) of Poltyrev [70] applied to fully random and binary block codes, and the lower bound on this minimal block length is calculated via the 1959 sphere-packing (SP59) bound of Shannon [89] and the improved sphere-packing (ISP) bound introduced in Section 5.3. In addition to bounds, this tradeoff between the block length (delay) and gap to capacity, is shown for some efficiently decodable error-correcting codes; the codes are taken from [108] (code 1), [26] (codes 2 and 4) and [25] (code 3).

than the TSB; this is expected since the error exponent of the TSB is slightly looser than the random-coding error exponent (see [104, upper plot of Fig. 3]). However, for small to moderate block lengths (i.e., for block lengths below approximately 5000 bits according to Figure 5.9), the TSB provides a tighter upper bound on the achievable gap as compared to the RCB. The improvement of the TSB over the RCB closes the gap between the upper and lower bounds on the achievable gap to capacity for small to moderate block lengths (where the lower bound is obtained via the SP59 bound which is tighter than the ISP bound for the considered range of block lengths). As for particularly efficient block codes, the code labeled 1 in Figure 5.9 is a block code of length 192 bits which is decoded using a near-ML decoder by applying 'box and match' decoding techniques [108]. It is observed that this code outperforms RCB for ML decoded random codes with the same block length and code rate, and almost coincides with the upper bound obtained via the TSB. It is also observed that this code achieves a block error probability of 10^{-5} at a gap to capacity of 2.76 dB while the SP59 bound gives that the block length required to achieve this performance is lower bounded by 133 bits (so the bound is very informative). The codes labeled 2, 3 and 4 are prototype-based LDPC codes of lengths 2048, 5176 and 8192 bits, respectively (codes 2 and 4 are taken from [26] and code 3 is taken from [25]). These codes achieve under iterative decoding a block error probability of 10^{-5} at gaps to capacity of 1.70, 1.27 and 1.07 dB, respectively. In terms of block length, the gap between the performance of these codes under iterative decoding and the SP59 lower bound on the block length required to achieve a block error probability of 10^{-5} at these channel conditions is less than one order of magnitude. It is also noted that throughout the range of block lengths depicted in Figure 5.9, the gap between the lower bound on the block length of optimal codes which stems from the better of the two sphere-packing bounds and the upper bound on the block length of random codes is less than one order of magnitude. This exemplifies the tightness of the sphere-packing bounds when used as lower bounds on the block lengths of optimal codes.

Figure 5.10 considers some LDPC codes of rate 0.88 bits per channel use which are BPSK modulated and transmitted over the AWGN channel. The gap to capacity in dB for which these codes achieve block error probabilities of 10^{-4} and 10^{-5} under iterative decoding is plotted as a function of block length. As in Figure 5.9, the figure uses upper and lower bounds on the achievable gap to capacity in terms of the block length: for this (relatively high) code rate and the considered range of block lengths, the ISP bound is uniformly tighter than the SP59 bound (so only the ISP bound is depicted in this figure, and the SP59 bound is omitted). The upper bounds on the required block lengths for achieving a target block error probability

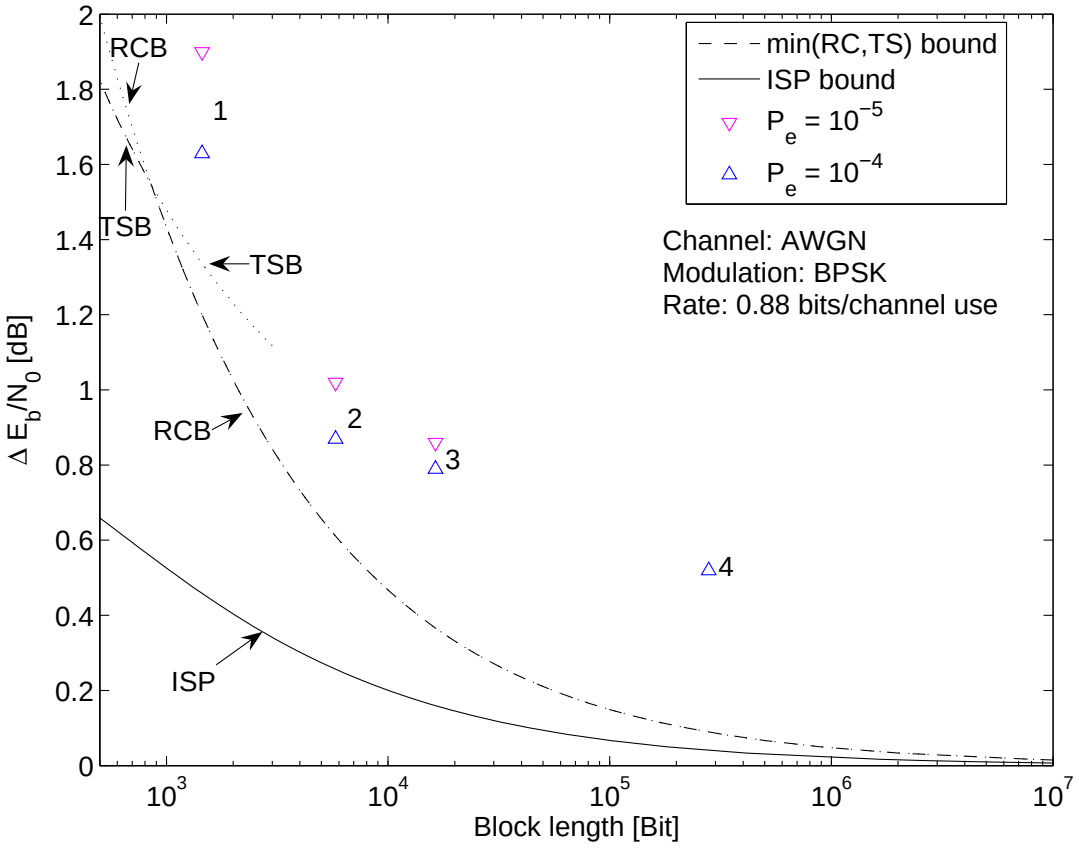

Figure 5.10: This figure refers to the tradeoff between the block length and the gap to capacity of error-correcting codes which are BPSK modulated and transmitted over an AWGN channel. The horizontal axis refers to the block length of the codes, and the vertical axis refers to the gap, measured in decibels, between the channel capacity and the energy per bit to spectral noise density ($\frac{E_\mathrm{b}}{N_0}$) which is required to obtain a block error probability P_e (set either to 10^{-4} or 10^{-5}). The considered rate of all the codes is 0.88 bit per channel use. The minimal gap to capacity which is required for achieving a block error probability of 10^{-5} is depicted via bounds: the upper bound is calculated using the random-coding bound (RCB) of Gallager [31] and the tangential-sphere bound (TSB) of Poltyrev [70] applied to fully random and binary block codes, and the lower bound on this minimal block length is calculated via the improved sphere-packing (ISP) bound introduced in Section 5.3 (which is better than the 1959 sphere-packing (SP59) bound of Shannon [89] for the considered code rate and the range of block lengths which is depicted in the horizontal line). In addition to bounds, this tradeoff between the block length (delay) and gap to capacity, is shown for some efficiently decodable error-correcting codes; the codes labeled by 1, 2, 3 and 4 are taken from [5], [25], [97] and [99], respectively.

in terms of the achievable gap to capacity are obtained via the RCB and the TSB when it is applied to the ensemble of fully random block codes. The upper and lower bounds refer to a block error probability of 10^{-5}. Similarly to Figure 5.9, the RCB is advantageous over the TSB for block codes of short to moderate block lengths; in this case, the advantage of the RCB over the TSB occurs for block lengths above approximately 1000 bits (instead of 5000 bits, as was the case in Fig. 5.9 for code rate of 0.5 bit per channel use). This shows that for short block lengths, the TSB is tighter than the RCB; however, since by increasing the code rate, the error exponent of the TSB becomes less tight as compared to the error exponent of the RCB (see the upper and lower plots of [104, Fig. 3] which refers to code rates of 0.5 and 0.9 bits per channel use), the asymptotic advantage of the RCB over the TSB is more pronounced, and the former bound is tighter than the latter already for shorter block lengths. The tradeoff between the gap to capacity (in terms of $\frac{E_b}{N_0}$) versus the block length is depicted in Figure 5.10 for some efficient error-correcting codes, in order to compare their practical performance and delay to the information-theoretic bounds (similarly to Fig. 5.9). For the examined block error probabilities (of 10^{-4} and 10^{-5}), the depicted codes require a gap to capacity of between 0.63 and 1.9 dB. For this range of $\frac{E_b}{N_0}$, the lower bound on the block lengths which is derived from the ISP bound is looser than the one given by the SP59 bound. However, both bounds are not very informative in this range. For cases where the gap to capacity is below 0.5 dB, the difference between the lower bound on the block length of optimal codes which stems from the ISP bound and the upper bound on the block length of random codes is less than one order of magnitude. Code number 1 is an LDPC of length 1448 bits whose construction of is based on balanced incomplete block designs [5]. This code achieves a block error probability of 10^{-5} at a gap to capacity of 1.9 dB while the RCB shows that the block length which is required to achieve this performance using random codes is upper bounded by 600 bits. The code labeled 2 is a prototype-based LDPC code of length 5176 bits which is taken from [25]. Code number 3 is a quasi-cyclic LDPC code of length 16362 bits taken from [97]. These code achieve under iterative decoding a block error probability of 10^{-5} at gaps to capacity of 1.02 and 0.86 dB, respectively. In terms of block length, the gap between the performance of these codes under iterative decoding and the upper bound on the block length of random codes which achieve a block error probability of 10^{-5} under the same channel conditions is less than one order of magnitude. The code labeled 4 is a finite-geometry LDPC code of length 279552 bits which is taken from [99]. For this code we only have the gap to capacity required to achieve a block error probability of 10^{-4}, however, it is clear that the difference in block length from the RCB becomes quite large as the gap to

capacity is reduced.

By fixing the block length and considering the gap in $\Delta E_{\mathrm{b}}/N_0$ between the performance of the specific codes and the sphere-packing bounds in Figures 5.9 and 5.10, it is observed that the codes considered in these plots exhibit gaps of 0.2–0.8 dB w.r.t. the information-theoretic limitation provided by the sphere-packing bounds (with the exception of code 1 in Figure 5.10 which exhibits a gap of about 1.25 dB). In this respect we also mention that some high rate turbo-product codes with moderate block lengths (see [21]) exhibit a gap of 0.75 – 0.95 dB w.r.t. the information-theoretic limitation provided by the ISP bound. Based on numerical results in [105] for the ensemble of uniformly interleaved (1144, 1000) turbo-block codes whose components are random systematic, binary and linear block codes, the gap in $\frac{E_{\mathrm{b}}}{N_0}$ between the ISP lower bound and an upper bound under ML decoding is 0.9 dB for a block error probability of 10^{-7}. These results exemplify the strength of the sphere-packing bounds for assessing the theoretical limitations of block codes and the power of iteratively decoded codes (see also [27, 44, 45, 79, 109]).

5.6 Summary

This paper presents an improved sphere-packing (ISP) bound for finite-length block codes whose transmission takes place over *symmetric* memoryless channels. The improved tightness of the bound is especially pronounced for codes of short to moderate block lengths, and some of its applications are exemplified in this paper. The derivation of the ISP bound was stimulated by the remarkable performance and feasible complexity of turbo-like codes with short to moderate block lengths. We were motivated by recent improvements on the sphere-packing bound of [87] for finite block lengths, as suggested by Valembois and Fossorier [109].

We first review the classical sphere-packing bounds, i.e., the 1959 sphere-packing bound (SP59) derived by Shannon for equal-energy signals transmitted over the Gaussian channel [89], and the 1967 sphere-packing (SP67) bound derived by Shannon, Gallager and Berlekamp for discrete memoryless channels [87]. The ISP bound, introduced in Section 5.3, is uniformly tighter than the classical SP67 bound [87] and the bound in [109].

We apply the ISP bound to various memoryless symmetric channels. The tightness of the ISP bound is exemplified by comparing it with upper and lower bounds on the ML decoding error probability and also with reported computer simulations of turbo-like codes under iterative decoding.

This paper also presents a new numerical algorithm which performs the entire

calculation of the SP59 bound in the logarithmic domain, thus facilitating the exact calculation of the SP59 bound for all block lengths without the need for asymptotic approximations. It is shown that the ISP bound suggests an interesting alternative to the SP59 bound, where the latter is specialized for the AWGN channel.

In a wide range of applications, one wishes to design a block code which satisfies a known delay constraint (i.e., the block length is limited) while adhering to a required performance over a given channel model. By fixing the communication channel model, code rate and the block error probability, sphere-packing bounds are transformed into lower bounds on the minimal block length required to achieve the target block error probability at a certain gap to capacity when an arbitrary block code and decoding algorithm are used. Comparing the performance of specific codes and decoding algorithms to the information-theoretic limitations provided by the sphere-packing bounds, enables one to deduce how far in terms of delay is a practical system from the fundamental limitations of information theory. Further details on the comparison between practically decodable codes and the sphere-packing bounds are found in Section 5.5.4.

The ISP bound is especially attractive for block codes of short to moderate block lengths, and its advantage is especially pronounced for high rate codes. Its improvement over the SP67 bound and the bound in [109, Theorem 7] also becomes more significant as the input alphabet of the considered modulation is increased.

Appendices

5.A Proof of Lemma 5.1

We consider a symmetric DMC with input alphabet $\mathcal{K} = \{0, \ldots, K-1\}$, output alphabet $\mathcal{J} = \{0, \ldots J-1\}$ (where $J, K \in \mathbb{N}$) and a transition probability function $P(\cdot|\cdot)$. Let $\{g_k\}_{k=0}^{K}$ be the set of unitary functions which satisfy the conditions (5.24) and (5.25) in Definition 5.2. To prove Lemma 5.1, we start with a discussion on the distribution $\mathbf{q}_s$ which satisfies (5.48).

On the input distribution $\mathbf{q}_s$ for symmetric DMCs

Lemma 5.A.1 For symmetric DMCs and an arbitrary value of $s \in (0, 1)$, the uniform distribution $q_{k,s} = \frac{1}{K}$ for $k \in \mathcal{K}$ satisfies (5.48) with equality.

Proof: To prove the lemma, it is required to show that

$$\sum_{j=0}^{J-1}\left\{P(j|k)^{1-s}\left(\sum_{k'=0}^{K-1}\frac{1}{K}P(j|k')^{1-s}\right)^{\frac{s}{1-s}}\right\}$$
$$=\sum_{j=0}^{J-1}\left(\sum_{k'=0}^{K-1}\frac{1}{K}P(j|k')^{1-s}\right)^{\frac{1}{1-s}}, \qquad \forall k\in\mathcal{K}. \tag{5.A.1}$$

Let us consider some $k\in\mathcal{K}$. Examining the left-hand side (LHS) of (5.A.1) gives

$$\sum_{j=0}^{J-1}\left\{P(j|k)^{1-s}\left(\sum_{k'=0}^{K-1}\frac{1}{K}P(j|k')^{1-s}\right)^{\frac{s}{1-s}}\right\}$$
$$=K\sum_{j=0}^{J-1}\left\{\frac{1}{K}P(j|k)^{1-s}\left(\sum_{k'=0}^{K-1}\frac{1}{K}P(j|k')^{1-s}\right)^{\frac{s}{1-s}}\right\}$$
$$\overset{(a)}{=}\sum_{\widetilde{k}=0}^{K-1}\sum_{j=0}^{J-1}\left\{\frac{1}{K}P(j|k)^{1-s}\left(\sum_{k'=0}^{K-1}\frac{1}{K}P(j|k')^{1-s}\right)^{\frac{s}{1-s}}\right\}$$
$$\overset{(b)}{=}\sum_{\widetilde{k}=0}^{K-1}\sum_{j=0}^{J-1}\left\{\frac{1}{K}P(g_{\widetilde{k}}(j)|k)^{1-s}\left(\sum_{k'=0}^{K-1}\frac{1}{K}P(g_{\widetilde{k}}(j)|k')^{1-s}\right)^{\frac{s}{1-s}}\right\} \tag{5.A.2}$$

where (a) holds by summing over a dummy variable $\widetilde{k}\in\mathcal{K}$ instead of the multiplication by K in the previous line, and (b) holds since $g_{\widetilde{k}}$ is unitary for all $\widetilde{k}\in\mathcal{K}$ (see (5.23) where the integral is replaced here by a sum). For all $j\in\mathcal{J}$ and $\widetilde{k}\in\mathcal{K}$, the symmetry properties in (5.24) - (5.26) give

$$\begin{aligned}P\left(g_{\widetilde{k}}(j)|\,k\right)&\overset{(a)}{=}P\left(\left(g_k^{-1}\circ g_{\widetilde{k}}\right)(j)|\,0\right)\\&\overset{(b)}{=}P\big(g_{(\widetilde{k}-k)\mathrm{mod}K}(j)|\,0\big)\\&\overset{(c)}{=}P\big(j|\,(k-\widetilde{k})\mathrm{mod}K\big)\end{aligned} \tag{5.A.3}$$

where (a) follows from (5.24), (b) relies on (5.25), and (c) follows from (5.24) and (5.26). Similarly, for all $j,\widetilde{k}\in\{0,1,\ldots,K-1\}$

$$\begin{aligned}\sum_{k'=0}^{K-1}\frac{1}{K}P(g_{\widetilde{k}}(j)|k')^{1-s}&\overset{(a)}{=}\sum_{k'=0}^{K-1}\frac{1}{K}P\big(j|\,(k'-\widetilde{k})\mathrm{mod}K\big)^{1-s}\\&\overset{(b)}{=}\sum_{k'=0}^{K-1}\frac{1}{K}P(j|k')^{1-s}\end{aligned} \tag{5.A.4}$$

where (a) relies on (5.A.3) and (b) holds since when the index k' takes all the values in $\{0,1,\ldots,K-1\}$, so does $(k'-\widetilde{k})\mathrm{mod}K$. Substituting (5.A.3) and (5.A.4) in (5.A.2)

gives

$$\sum_{j=0}^{J-1}\left\{P(j|k)^{1-s}\left(\sum_{k'=0}^{K-1}\frac{1}{K}P(j|k')^{1-s}\right)^{\frac{s}{1-s}}\right\}$$

$$=\sum_{\widetilde{k}=0}^{K-1}\sum_{j=0}^{J-1}\left\{\frac{1}{K}P\big(j|(k-\widetilde{k})\mathrm{mod}K\big)^{1-s}\left(\sum_{k'=0}^{K-1}\frac{1}{K}P(j|k')^{1-s}\right)^{\frac{s}{1-s}}\right\}$$

$$=\sum_{j=0}^{J-1}\left\{\left(\sum_{\widetilde{k}=0}^{K-1}\frac{1}{K}P\big(j|(k-\widetilde{k})\mathrm{mod}K\big)^{1-s}\right)\left(\sum_{k'=0}^{K-1}\frac{1}{K}P(j|k')^{1-s}\right)^{\frac{s}{1-s}}\right\}$$

$$\overset{(a)}{=}\sum_{j=0}^{J-1}\left\{\left(\sum_{\widetilde{k}=0}^{K-1}\frac{1}{K}P(j|\widetilde{k})^{1-s}\right)\left(\sum_{k'=0}^{K-1}\frac{1}{K}P(j|k')^{1-s}\right)^{\frac{s}{1-s}}\right\}$$

$$=\sum_{j=0}^{J-1}\left(\sum_{k'=0}^{K-1}\frac{1}{K}P(j|k')^{1-s}\right)^{\frac{1}{1-s}} \tag{5.A.5}$$

where equality (a) holds since the $\widetilde{k}$ takes all the values in $\{0,1,\ldots,K-1\}$, and so does the index $k' \triangleq (k-\widetilde{k})\mathrm{mod}K$. ■

We now turn to explore how the symmetry of the channel and the input distribution $\mathbf{q}_s$ induce a symmetry on the probability tilting measure f_s.

On the symmetry of the tilting measure f_s for strictly symmetric DMCs

Lemma 5.A.2 For all $s \in (0,1)$, $k \in \mathcal{K}$ and $j \in \mathcal{J}$, the tilting measure f_s in (5.50) satisfies

$$f_s(j) = f_s\big(g_k(j)\big). \tag{5.A.6}$$

Proof: Examining the definition of f_s in (5.50), it can be observed that it suffices to show that

$$\alpha_{j,s} = \alpha_{g_k(j),s}\,, \quad \forall s \in (0,1),\, k \in \mathcal{K},\, j \in \mathcal{J} \tag{5.A.7}$$

where $\alpha_{j,s}$ is given in (5.49). Note that for the uniform input distribution where $q_{k,s} = \frac{1}{K}$ for all $k \in \mathcal{K}$, inequalities (5.48) and (5.49) hold with equality (see Lemma 5.A.1). From (5.A.4), equality (5.A.7) follows for the uniform input distribution. ■

Having established some symmetry properties of $\mathbf{q}_s$ and f_s, we are ready to prove equalities (5.51) – (5.53).

On the independence of μ_k and its two derivatives from k As we have shown, the uniform distribution $\mathbf{q}_s$ satisfies (5.48) in equality for all inputs, so

$$\begin{aligned}
\mu_k(s) &= \ln\left(\sum_j P(j|k)^{1-s} f_s(j)^s\right) \\
&\overset{(a)}{=} \ln\left(\sum_j P(j|k)^{1-s} (\alpha_{j,s})^{\frac{s}{1-s}}\right) - s\ln\left(\sum_j (\alpha_{j,s})^{\frac{1}{1-s}}\right) \\
&\overset{(b)}{=} (1-s)\ln\left(\sum_j (\alpha_{j,s})^{\frac{1}{1-s}}\right) \\
&\overset{(c)}{=} (1-s)\ln\left(\sum_j \left[\sum_k q_{k,s} P(j|k)^{1-s}\right]^{\frac{1}{1-s}}\right) \qquad (5.A.8)
\end{aligned}$$

where (a) follows from the choice of f_s in (5.49) and (5.50), (b) follows from Lemma 5.A.1 and (5.49), and (c) follows from (5.49). Under the setting $s = \frac{\rho}{1+\rho}$, since the conditions on $\mathbf{q_s}$ in (5.48) are identical to the conditions on the input distribution $\mathbf{q} = \mathbf{q_s}$ which maximizes $E_0(\frac{s}{1-s}, \mathbf{q})$ as stated in [31, Theorem 4], then

$$\begin{aligned}
\mu_k(s, f_s) &= (1-s)\ln\left(\sum_j \left[\sum_k q_{k,s} P(j|k)^{\frac{1}{1+\frac{s}{1-s}}}\right]^{1+\frac{s}{1-s}}\right) \\
&= -(1-s)\, E_0\left(\frac{s}{1-s}, \mathbf{q_s}\right) \\
&= -(1-s)\, E_0\left(\frac{s}{1-s}\right), \quad 0 < s < 1 \qquad (5.A.9)
\end{aligned}$$

where E_0 is given in (5.19). This proves (5.51).

We now turn to prove the independence of the first two derivatives of μ_k w.r.t s from $k \in \mathcal{K}$.

Remark 5.11 Note that the partial derivative of $\mu_k(s)$ w.r.t s is performed while holding $f = f_s$ constant.

As is shown in [87],

$$\begin{aligned}
\mu'(s) &= \mathbb{E}_{Q_s}\big(D(j)\big) \\
\mu''(s) &= \mathrm{Var}_{Q_s}\big(D(j)\big)
\end{aligned}$$

where

$$D(j) \triangleq \ln\left(\frac{P_2(j)}{P_1(j)}\right), \quad Q_s(j) \triangleq \frac{P_1(j)^{1-s} P_2(j)^s}{\sum_{j'} P_1(j')^{1-s} P_2(j')^s}.$$

For every $k \in \mathcal{K}$, P_1 and P_2 used in μ_k are defined to be $P(\cdot|k)$ and f_s, respectively. Hence, for all $k \in \mathcal{K}$

$$\begin{aligned} \mu_k'(s, f_s) &= \mathbb{E}_{Q_{k,s}}\big(D_{k,s}(j)\big) \\ \mu_k''(s, f_s) &= \mathrm{Var}_{Q_{k,s}}\big(D_{k,s}(j)\big) \end{aligned} \tag{5.A.10}$$

where

$$D_{k,s}(j) \triangleq \ln\left(\frac{f_s(j)}{P(j|k)}\right) \tag{5.A.11}$$

$$Q_{k,s}(j) \triangleq \frac{P(j|k)^{1-s} f_s(j)^s}{\sum\limits_{j'=0}^{J-1} P(j'|k)^{1-s} f_s(j')^s}. \tag{5.A.12}$$

Applying (5.24) and Lemma 5.A.2, we get that for all $k \in \mathcal{K}$

$$\begin{aligned} \sum_{j'=0}^{J-1} P(j'|k)^{1-s} f_s(j')^s &\overset{(a)}{=} \sum_{j'=0}^{J-1} P(g_k^{-1}(j')|0)^{1-s} f_s(g_k^{-1}(j'))^s \\ &\overset{(b)}{=} \sum_{j'=0}^{J-1} P(j'|0)^{1-s} f_s(j')^s \end{aligned} \tag{5.A.13}$$

where (a) follows from (5.24) and Lemma 5.A.2, and (b) follows since g_k^{-1} is unitary. Substituting (5.A.13) in (5.A.12) gives

$$\begin{aligned} Q_{k,s}(j) &= \frac{P(j|k)^{1-s} f_s(j)^s}{\sum\limits_{j'=0}^{J-1} P(j'|k)^{1-s} f_s(j')^s} \\ &\overset{(a)}{=} \frac{P(g_k^{-1}(j)|0)^{1-s} f_s(g_k^{-1}(j))^s}{\sum\limits_{j'=0}^{J-1} P(j'|0)^{1-s} f_s(j')^s} \\ &\overset{(b)}{=} Q_{0,s}\big(g_k^{-1}(j)\big) \end{aligned} \tag{5.A.14}$$

where (a) follows from (5.24), (5.26), (5.50), (5.A.6) and (5.A.13), and (b) relies on the definition of $Q_{k,s}$ in (5.A.12). Similarly,

$$\begin{aligned} D_{k,s}(j) &= \ln\left(\frac{f_s(j)}{P(j|k)}\right) \\ &\overset{(a)}{=} \ln\left(\frac{f_s\big(g_k^{-1}(j)\big)}{P\big(g_k^{-1}(j)|0\big)}\right) \\ &\overset{(b)}{=} D_{0,s}\big(g_k^{-1}(j)\big) \end{aligned} \tag{5.A.15}$$

where (a) follows from (5.24), (5.26) and (5.A.6), and (b) relies on the definition of $D_{k,s}$ in (5.A.11). Using (5.A.14) and (5.A.15), we finally get for all $k \in \mathcal{K}$

$$\begin{aligned}
\mu_k'(s) &= \mathbb{E}_{Q_{k,s}}\big(D_{k,s}(j)\big) \\
&= \sum_{j=0}^{J-1} Q_{k,s}(j) D_{k,s}(j) \\
&= \sum_{j=0}^{J-1} Q_{0,s}\big(g_k^{-1}(j)\big) D_{0,s}\big(g_k^{-1}(j)\big) \\
&\overset{(a)}{=} \sum_{j=0}^{J-1} Q_{0,s}(j) D_{0,s}(j) \\
&= \mu_0'(s)
\end{aligned}$$

and

$$\begin{aligned}
\mu_k''(s) &= \mathrm{Var}_{Q_{k,s}}\big(D_{k,s}(j)\big) \\
&= \sum_{j=0}^{J-1} Q_{k,s}(j) D_{k,s}^2(j) - \mu_k'(s)^2 \\
&= \sum_{j=0}^{J-1} Q_{0,s}\big(g_k^{-1}(j)\big) \Big(D_{0,s}\big(g_k^{-1}(j)\big)\Big)^2 - \mu_0'(s)^2 \\
&\overset{(b)}{=} \sum_{j=0}^{J-1} Q_{0,s}(j) \big(D_{0,s}(j)\big)^2 - \mu_0'(s)^2 \\
&= \mu_0''(s)
\end{aligned}$$

where (a) and (b) follow since g_k^{-1} is unitary for all $k \in \mathcal{K}$. This completes the proof of Lemma 5.1.

Remark 5.12 Equalities (5.51)–(5.53) hold for arbitrary symmetric memoryless channels. For a general output alphabet $\mathcal{J} \subseteq \mathbb{R}^d$, the proof of these properties follows the same lines as the proof here with the exception that the sums over $\mathcal{J}$ are replaced by integrals. As in Definition 5.1, if the projection of $\mathcal{J}$ over some of the d dimensions is countable, the integration over these dimensions is turned into a sum.

5.B Calculation of the Function μ_0 in (5.47) for some Symmetric Channels

This appendix presents some technical calculations which yield the expressions for the function μ_0 defined in (5.47) and its first two derivatives w.r.t. s (while holding f_s fixed in the calculation of the partial derivatives of μ w.r.t. s, as required in [87]). The examined cases are M-ary PSK modulated signals transmitted over fully interleaved fading channels, with the AWGN channel as a special case, and binary block codes transmitted over the BEC. These expressions serve for the application of the VF bound in [109] and the ISP bound derived in Section 5.3 to block codes transmitted over these channels.

5.B.1 M-ary PSK Modulated Signal over Fully Interleaved Fading Channels with Perfect CSI

For M-ary PSK modulated signals transmitted over a fully interleaved fading channel, the channel output is $\mathcal{J} = \mathbb{R}^2 \times \mathbb{R}^+$, where the first two coordinates refer to the vector $\mathbf{Y}$ and the third refers to the fading coefficient A. In the case of a continuous output alphabet, the sums in (5.A.8) are replaced by integrals, and the transition probabilities are replaced by transition probability density functions. To simplify the presentation, for all $s \in (0,1)$, $\mathbf{y} \in \mathbb{R}^2$ and $a \in \mathbb{R}^+$ we define

$$\kappa_s(\mathbf{y}, a) \triangleq \left(\frac{1}{M} \sum_{k=0}^{M-1} e^{\frac{(1-s)a \, \langle \mathbf{y}, \mathbf{x}_k - \mathbf{x}_0 \rangle}{\sigma^2}} \right)^{\frac{1}{1-s}} . \tag{5.B.1}$$

This expression will be used throughout the following calculations.

Due to the symmetry of the channel, we get from Lemma 5.A.1 that the distribution $\mathbf{q}_s$ which satisfies (5.48) is uniform. Hence, we get by substituting (5.80) into (5.A.8) that

$$\mu_0(s) = (1-s) \ln \left(\iint_{\mathbb{R}^2} \int_0^\infty \frac{p_A(a)}{2\pi\sigma^2} \, e^{-\frac{\|\mathbf{y} - a\mathbf{x}_0\|^2}{2\sigma^2}} \zeta_s(\mathbf{y}, a) \mathrm{d}a \, \mathrm{d}\mathbf{y} \right)$$

where

$$\zeta_s(\mathbf{y}, a) \triangleq \left(\frac{1}{M} \sum_{k=0}^{M-1} e^{-\frac{(1-s)(\|\mathbf{y} - a\mathbf{x}_k\|^2 - \|\mathbf{y} - a\mathbf{x}_0\|^2)}{2\sigma^2}} \right)^{\frac{1}{1-s}}$$

Since $\|x_k\|^2 = 1$ for all $k \in \{0, 1, \ldots, M-1\}$ we have

$$\|\mathbf{y} - a\mathbf{x}_k\|^2 - \|\mathbf{y} - a\mathbf{x}_0\|^2 = -2a \langle \mathbf{y}, \mathbf{x}_k - \mathbf{x}_0 \rangle \tag{5.B.2}$$

and so μ_0 can be rewritten in the form

$$\mu_0(s) = (1-s)\ln\left(\theta(s)\right)$$

where

$$\theta(s) \triangleq \iint\limits_{\mathbb{R}^2} \int\limits_0^\infty \frac{p_A(a)}{2\pi\sigma^2} e^{-\frac{\|\mathbf{y}-a\mathbf{x}_0\|^2}{2\sigma^2}} \kappa_s(\mathbf{y},a)\mathrm{d}a\,\mathrm{d}\mathbf{y}. \tag{5.B.3}$$

and $\kappa_s(\mathbf{y},a)$ is defined in (5.B.1).

We now turn to calculate the derivative of μ_0 with respect to s while holding $f = f_s$ constant. Substituting (5.80) into the definition of f_s in (5.50), we get that f_s is given by

$$\begin{aligned} f_s(\mathbf{y},a) &= \frac{\left(\sum\limits_{k=0}^{M-1} \frac{1}{M}\left(\frac{p_A(a)}{2\pi\sigma^2}\right)^{1-s} e^{-\frac{(1-s)\,\|\mathbf{y}-a\mathbf{x}_k\|^2}{2\sigma^2}}\right)^{\frac{1}{1-s}}}{\iint\limits_{\mathbb{R}^2}\int\limits_0^\infty \left(\sum\limits_{k=0}^{M-1} \frac{1}{M}\left(\frac{p_A(a')}{2\pi\sigma^2}\right)^{1-s} e^{-\frac{(1-s)\,\|\mathbf{y}'-a'\mathbf{x}_k\|^2}{2\sigma^2}}\right)^{\frac{1}{1-s}} \mathrm{d}a'\mathrm{d}\mathbf{y}'} \\ &= \frac{\kappa_s(\mathbf{y},a)}{\theta(s)}\,\frac{p_A(a)}{2\pi\sigma^2}\,e^{-\frac{\|\mathbf{y}-a\mathbf{x}_0\|^2}{2\sigma^2}} \end{aligned} \tag{5.B.4}$$

where the last equality follows from (5.B.2) and (5.B.3). The log-likelihood ratio $D_{0,s}$ in (5.A.11) is given by

$$\begin{aligned} D_{0,s}(\mathbf{y},a) &\triangleq \ln\left(\frac{f_s(\mathbf{y},a)}{P(\mathbf{y},a|0)}\right) \\ &= \ln\Big(\kappa_s(\mathbf{y},a)\Big) - \ln\Big(\theta(s)\Big) \end{aligned} \tag{5.B.5}$$

where the second equality follows from (5.80) and (5.B.4). The distribution $Q_{0,s}$ in (5.A.12) is given by

$$\begin{aligned} &Q_{0,s}(\mathbf{y},a) \\ &\triangleq \frac{P(\mathbf{y},a|0)^{1-s} f_s(\mathbf{y},a)^s}{\iint\limits_{\mathbb{R}^2}\int\limits_0^\infty P(\mathbf{y}',a'|0)^{1-s} f_s(\mathbf{y}',a')^s\,\mathrm{d}a'\,\mathrm{d}\mathbf{y}'} \\ &= \frac{\frac{p_A(a)}{2\pi\sigma^2} e^{-\frac{\|\mathbf{y}-a\mathbf{x}_0\|^2}{2\sigma^2}} \Big(\kappa_s(\mathbf{y},a)\Big)^s}{\iint\limits_{\mathbb{R}^2}\int\limits_0^\infty \frac{p_A(a')}{2\pi\sigma^2} e^{-\frac{\|\mathbf{y}'-a'\mathbf{x}_0\|^2}{2\sigma^2}} \Big(\kappa_s(\mathbf{y},a')\Big)^s\,\mathrm{d}a'\mathrm{d}\mathbf{y}'} \end{aligned}$$

$$
\begin{aligned}
&\overset{(a)}{=} \frac{p_A(a)}{2\pi\sigma^2} e^{-\frac{\|\mathbf{y}-a\mathbf{x}_0\|^2}{2\sigma^2}} \big(\kappa_s(\mathbf{y},a)\big)^s \\
&\quad \cdot \left[\iint\limits_{\mathbb{R}^2} \int\limits_0^\infty \left(\frac{1}{M} \sum_{k=0}^{M-1} \left(\frac{1}{2\pi\sigma^2}\, e^{-\frac{\|\mathbf{y}'-a'\mathbf{x}_k\|^2}{2\sigma^2}} \right)^{1-s} \right)^{\frac{s}{1-s}} \left(\frac{p_A(a')}{2\pi\sigma^2}\, e^{-\frac{\|\mathbf{y}'-a'\mathbf{x}_0\|^2}{2\sigma^2}} \right)^{1-s} \mathrm{d}a'\, \mathrm{d}\mathbf{y}' \right]^{-1} \\
&\overset{(b)}{=} \frac{\frac{p_A(a)}{2\pi\sigma^2} e^{-\frac{\|\mathbf{y}-a\mathbf{x}_0\|^2}{2\sigma^2}} \big(\kappa_s(\mathbf{y},a)\big)^s}{\iint\limits_{\mathbb{R}^2} \int\limits_0^\infty \left(\frac{1}{M} \sum_{k=0}^{M-1} \left(\frac{p_A(a')}{2\pi\sigma^2} e^{-\frac{\|\mathbf{y}'-a'\mathbf{x}_k\|^2}{2\sigma^2}} \right)^{1-s} \right)^{\frac{1}{1-s}} \mathrm{d}a'\mathrm{d}\mathbf{y}'} \\
&\overset{(c)}{=} \frac{\frac{p_A(a)}{2\pi\sigma^2} e^{-\frac{\|\mathbf{y}-a\mathbf{x}_0\|^2}{2\sigma^2}} \big(\kappa_s(\mathbf{y},a)\big)^s}{\iint\limits_{\mathbb{R}^2} \int\limits_0^\infty \frac{p_A(a')}{2\pi\sigma^2} e^{-\frac{\|\mathbf{y}'-a'\mathbf{x}_0\|^2}{2\sigma^2}} \kappa_s(\mathbf{y},a')\, \mathrm{d}a'\mathrm{d}\mathbf{y}'} \\
&\overset{(d)}{=} \frac{p_A(a)\,\big(\kappa_s(\mathbf{y},a)\big)^s}{2\pi\sigma^2\,\theta(s)}\, e^{-\frac{\|\mathbf{y}-a\mathbf{x}_0\|^2}{2\sigma^2}}
\end{aligned}
\tag{5.B.6}
$$

where (a) and (c) rely on (5.B.2), (b) follows from Lemma 2.1 in the proof for symmetric memoryless channels, and (d) relies on the definition of θ in (5.B.3). Substituting (5.B.5) and (5.B.6) in (5.A.10) we get

$$
\begin{aligned}
\mu_0'(s) &= \mathbb{E}_{Q_{0,s}}\big(D_{0,s}\big) \\
&= \frac{1}{\theta(s)} \iint\limits_{\mathbb{R}^2} \int\limits_0^\infty \frac{p_A(a)}{2\pi\sigma^2} e^{-\frac{\|\mathbf{y}-a\mathbf{x}_0\|^2}{2\sigma^2}} \big(\kappa_s(\mathbf{y},a)\big)^s \\
&\qquad\qquad \cdot \ln\big(\kappa_s(\mathbf{y},a)\big)\, \mathrm{d}a\, \mathrm{d}\mathbf{y} - \ln\big(\theta(s)\big)
\end{aligned}
\tag{5.B.7}
$$

and

$$
\begin{aligned}
\mu_0''(s) &= \mathbb{E}_{Q_{0,s}}\big(D_{0,s}^2(\mathbf{y})\big) - \mu_0'(s)^2 \\
&= \frac{1}{\theta(s)} \iint\limits_{\mathbb{R}^2} \int\limits_0^\infty \frac{p_A(a)}{2\pi\sigma^2} e^{-\frac{\|\mathbf{y}-a\mathbf{x}_0\|^2}{2\sigma^2}} \big(\kappa_s(\mathbf{y},a)\big)^s \\
&\qquad\qquad \cdot \Big(\ln\big(\kappa_s(\mathbf{y},a)\big) - \ln\big(\theta(s)\big) \Big)^2 \mathrm{d}a\, \mathrm{d}\mathbf{y} \\
&\qquad -\mu_0'(s)^2 .
\end{aligned}
\tag{5.B.8}
$$

In this paper, we consider the particular case where the fading coefficients have a Rayleigh distribution. In this case, the distribution of the fading samples is given by $p_A(a) = 2a\, e^{-a^2}$ for $a \geq 0$, so that $\mathbb{E}(A^2) = 1$.

5.B.2 M-ary PSK Modulated Signals over the AWGN Channel

A widely studied special case of fully interleaved fading channels is the AWGN channel where the fading coefficients are set to 1. Substituting $P_A(a) = \delta(a-1)$, where δ is the Dirac delta function at zero, we get that θ in (5.B.3) is particularized to

$$\theta(s) \triangleq \iint\limits_{\mathbb{R}^2} \frac{e^{-\frac{\|\mathbf{y}-\mathbf{x}_0\|^2}{2\sigma^2}}}{2\pi\sigma^2}\, \kappa_s(\mathbf{y}, a)\, \mathrm{d}\mathbf{y} \tag{5.B.9}$$

and the first and second derivatives of μ_0 w.r.t. s (while holding f_s constant) are given by

$$\mu_0'(s) = \frac{1}{\theta(s)} \iint\limits_{\mathbb{R}^2} \frac{e^{-\frac{\|\mathbf{y}-\mathbf{x}_0\|^2}{2\sigma^2}}}{2\pi\sigma^2}\, \big(\kappa_s(\mathbf{y}, a)\big)^s\, \ln\big(\kappa_s(\mathbf{y}, a)\big)\, \mathrm{d}\mathbf{y} - \ln\big(\theta(s)\big) \tag{5.B.10}$$

and

$$\mu_0''(s) = \frac{1}{\theta(s)} \iint\limits_{\mathbb{R}^2} \frac{1}{2\pi\sigma^2} e^{-\frac{\|\mathbf{y}-\mathbf{x}_0\|^2}{2\sigma^2}} \big(\kappa_s(\mathbf{y}, a)\big)^s \Big(\ln\big(\kappa_s(\mathbf{y}, a)\big) - \ln\big(\theta(s)\big)\Big)^2 \mathrm{d}\mathbf{y} - \mu_0'(s)^2\,. \tag{5.B.11}$$

5.B.3 The Binary Erasure Channel

Let us denote the output of the channel when an erasure has occurred by $\mathcal{E}$, and let p designate the erasure probability of the channel. Since the BEC is symmetric, the input distribution $\mathbf{q}_s$ which satisfies (5.48) is uniform (see Lemma 5.A.1), and we get from (5.A.8)

$$\begin{aligned} \mu_0(s, f_s) &= (1-s)\ln\left(\frac{2(1-p)}{2^{\frac{1}{1-s}}} + p\right) \\ &= (1-s)\ln\left(2(1-p) + 2^{\frac{1}{1-s}} p\right) - \ln 2\,. \end{aligned} \tag{5.B.12}$$

We now turn to calculate f_s for the BEC; substituting the transition probabilities into (5.50) gives

$$\begin{aligned} f_s(0) = f_s(1) &= \frac{\left(\frac{1}{2}(1-p)^{1-s}\right)^{\frac{1}{1-s}}}{2\left(\frac{1}{2}(1-p)^{1-s}\right)^{\frac{1}{1-s}} + (p^{1-s})^{\frac{1}{1-s}}} \\ &= \frac{2^{-\frac{1}{1-s}}(1-p)}{2^{1-\frac{1}{1-s}}(1-p) + p} \\ &= \frac{1-p}{2(1-p) + 2^{\frac{1}{1-s}} p} \end{aligned} \tag{5.B.13}$$

and

$$\begin{aligned} f_s(\mathcal{E}) &= \frac{(p^{1-s})^{\frac{1}{1-s}}}{2\left(\frac{1}{2}(1-p)^{1-s}\right)^{\frac{1}{1-s}} + (p^{1-s})^{\frac{1}{1-s}}} \\ &= \frac{p}{2^{1-\frac{1}{1-s}}(1-p)+p} \\ &= \frac{2^{\frac{1}{1-s}}p}{2(1-p)+2^{\frac{1}{1-s}}p}. \end{aligned} \tag{5.B.14}$$

Substituting (5.B.13) and (5.B.14) into the definition of the distribution $Q_{0,s}$ in (5.A.12) gives

$$\begin{aligned} Q_{0,s}(0) &= \frac{P(0|0)^{1-s}f_s(0)^s}{\sum_{j\in\{0,1,\mathcal{E}\}} P(j|0)^{1-s}f_s(j)^s} = \frac{1-p}{1-p+2^{\frac{s}{1-s}}p} \\ Q_{0,s}(1) &= \frac{P(1|0)^{1-s}f_s(0)^s}{\sum_{j\in\{0,1,\mathcal{E}\}} P(j|0)^{1-s}f_s(j)^s} = 0 \\ Q_{0,s}(\mathcal{E}) &= \frac{P(\mathcal{E}|0)^{1-s}f_s(\mathcal{E})^s}{\sum_{j\in\{0,1,\mathcal{E}\}} P(j|0)^{1-s}f_s(j)^s} = \frac{2^{\frac{s}{1-s}}p}{1-p+2^{\frac{s}{1-s}}p} \end{aligned} \tag{5.B.15}$$

and the LLR in (5.A.11) is given by

$$\begin{aligned} D_{0,s}(0) &= \ln\left(\frac{1}{2(1-p)+2^{\frac{1}{1-s}}p}\right) \\ D_{0,s}(\mathcal{E}) &= \ln\left(\frac{2^{\frac{1}{1-s}}}{2(1-p)+2^{\frac{1}{1-s}}p}\right). \end{aligned} \tag{5.B.16}$$

Applying (5.B.15) and (5.B.16) we get from (5.A.12)

$$\begin{aligned} \mu_0'(s,f_s) &= \mathbb{E}_{Q_{0,s}}\big(D_{0,s}\big) \\ &= \frac{1-p}{1-p+2^{\frac{s}{1-s}}p}\ln\left(\frac{1}{2(1-p)+2^{\frac{1}{1-s}}p}\right) \\ &\quad + \frac{2^{\frac{s}{1-s}}p}{1-p+2^{\frac{s}{1-s}}p}\ln\left(\frac{2^{\frac{1}{1-s}}}{2(1-p)+2^{\frac{1}{1-s}}p}\right) \\ &= \ln\left(\frac{1}{1-p+2^{\frac{s}{1-s}}p}\right) + \frac{2^{\frac{s}{1-s}}p}{1-p+2^{\frac{s}{1-s}}p}\,\frac{\ln 2}{1-s} \end{aligned} \tag{5.B.17}$$

and

$$\begin{aligned}
\mu_0''(s,f_s) &= \mathbb{E}_{Q_{0,s}}\big(D_{0,s}^2\big) - \mu_0'(s,f_s)^2 \\
&= \frac{1-p}{1-p+2^{\frac{s}{1-s}}p}\ln^2\left(\frac{1}{2(1-p)+2^{\frac{1}{1-s}}p}\right) \\
&\quad + \frac{2^{\frac{s}{1-s}}p}{1-p+2^{\frac{s}{1-s}}p}\ln^2\left(\frac{2^{\frac{1}{1-s}}}{2(1-p)+2^{\frac{1}{1-s}}p}\right) - \mu_0'(s,f_s)^2 \\
&= \ln^2\left(\frac{1}{1-p+2^{\frac{s}{1-s}}p}\right) + \frac{2^{\frac{s}{1-s}}p}{1-p+2^{\frac{s}{1-s}}p}\left(\frac{\ln 2}{1-s}\right)^2 \\
&\quad + \frac{2^{\frac{1}{1-s}}p}{1-p+2^{\frac{s}{1-s}}p}\frac{\ln 2}{1-s}\ln\left(\frac{1}{1-p+2^{\frac{s}{1-s}}p}\right) - \mu_0'(s,f_s)^2 \\
&= \frac{2^{\frac{s}{1-s}}\,p\,(1-p)}{\big(1-p+2^{\frac{s}{1-s}}p\big)^2}\left(\frac{\ln 2}{1-s}\right)^2 \qquad (5.B.18)
\end{aligned}$$

5.C Proof of Proposition 5.3

From the definition of f_N in (5.67), it follows that

$$\begin{aligned}
f_N(x) &= \frac{1}{2^{\frac{N-1}{2}}\Gamma(\frac{N+1}{2})}\int_0^\infty z^{N-1}\exp(-\frac{z^2}{2}+zx)\,\mathrm{d}z \\
&= \frac{\exp\left(\frac{x^2}{2}\right)}{2^{\frac{N-1}{2}}\Gamma(\frac{N+1}{2})}\int_0^\infty z^{N-1}\exp\left(-\frac{(z-x)^2}{2}\right)\,\mathrm{d}z \\
&= \frac{\exp\left(\frac{x^2}{2}\right)}{2^{\frac{N-1}{2}}\Gamma(\frac{N+1}{2})}\int_{-x}^\infty (u+x)^{N-1}\exp\left(-\frac{u^2}{2}\right)\,\mathrm{d}u\,.
\end{aligned}$$

From the binomial formula, we get

$$f_N(x) = \frac{\exp\left(\frac{x^2}{2}\right)}{2^{\frac{N-1}{2}}\Gamma(\frac{N+1}{2})}\sum_{j=0}^{N-1}\left[\binom{N-1}{j}x^{N-1-j}\int_{-x}^\infty u^j\exp\left(-\frac{u^2}{2}\right)\,\mathrm{d}u\right]. \qquad (5.C.1)$$

We now examine the integrals on the RHS of (5.C.1). For odd values of j, we get

$$\begin{aligned}
\int_{-x}^\infty u^j\exp\left(-\frac{u^2}{2}\right)\mathrm{d}u &= \int_{-x}^x u^j\exp\left(-\frac{u^2}{2}\right)\mathrm{d}u + \int_x^\infty u^j\exp\left(-\frac{u^2}{2}\right)\mathrm{d}u \\
&= \int_x^\infty u^j\exp\left(-\frac{u^2}{2}\right)\mathrm{d}u \\
&= \int_0^\infty u^j\exp\left(-\frac{u^2}{2}\right)\mathrm{d}u - \int_0^x u^j\exp\left(-\frac{u^2}{2}\right)\mathrm{d}u \qquad (5.C.2)
\end{aligned}$$

where the second equality follows since the integrand is an odd function for odd values of j, and the interval of first integral is symmetric around zero (so this integral

vanishes). For even values of j, we get

$$\begin{aligned}\int_{-x}^{\infty} u^j \exp\left(-\frac{u^2}{2}\right) \mathrm{d}u &= \int_0^{\infty} u^j \exp\left(-\frac{u^2}{2}\right) \mathrm{d}u + \int_{-x}^{0} u^j \exp\left(-\frac{u^2}{2}\right) \mathrm{d}u \\ &= \int_0^{\infty} u^j \exp\left(-\frac{u^2}{2}\right) \mathrm{d}u + \int_0^{x} u^j \exp\left(-\frac{u^2}{2}\right) \mathrm{d}u \quad (5.C.3)\end{aligned}$$

where the second equality holds since the integrand is an even function for even values of j. Combining (5.C.2) and (5.C.3) gives that for $j \in \{0, 1, \ldots, N-1\}$

$$\begin{aligned}\int_{-x}^{\infty} u^j \exp\left(-\frac{u^2}{2}\right) \mathrm{d}u &= \int_0^{\infty} u^j \exp\left(-\frac{u^2}{2}\right) \mathrm{d}u + (-1)^j \int_0^{x} u^j \exp\left(-\frac{u^2}{2}\right) \mathrm{d}u \\ &\overset{(a)}{=} \int_0^{\infty} (2t)^{\frac{j-1}{2}} \exp(-t)\, \mathrm{d}t + (-1)^j \int_0^{\frac{x^2}{2}} (2t)^{\frac{j-1}{2}} \exp(-t)\, \mathrm{d}t \\ &= 2^{\frac{j-1}{2}} \int_0^{\infty} t^{\frac{j-1}{2}} \exp(-t)\, \mathrm{d}t \left[1 + (-1)^j \frac{\displaystyle\int_0^{\frac{x^2}{2}} t^{\frac{j-1}{2}} \exp(-t)\, \mathrm{d}t}{\displaystyle\int_0^{\infty} t^{\frac{j-1}{2}} \exp(-t)\, \mathrm{d}t}\right] \\ &= 2^{\frac{j-1}{2}} \Gamma\left(\frac{j+1}{2}\right) \left[1 + (-1)^j \tilde{\gamma}\left(\frac{x^2}{2}, \frac{j+1}{2}\right)\right]\end{aligned}$$

where (a) follows by substituting $t \triangleq \frac{u^2}{2}$ and the functions Γ and $\tilde{\gamma}$ are introduced in (5.75) and (5.76), respectively. Substituting the last equality in (5.C.1) and also noting that

$$\binom{N-1}{j} = \frac{\Gamma(N)}{\Gamma(N-j)\,\Gamma(j+1)}, \quad N \in \mathbb{N},\, j \in \{0, 1, \ldots, N-1\}$$

we get

$$\begin{aligned}f_N(x) &= \frac{\exp\left(\frac{x^2}{2}\right)}{2^{\frac{N-1}{2}}\Gamma(\frac{N+1}{2})} \sum_{j=0}^{N-1} \left\{ \frac{\Gamma(N)}{\Gamma(N-j)\,\Gamma(j+1)}\, x^{N-1-j}\, 2^{\frac{j-1}{2}} \right. \\ &\qquad \left. \cdot \Gamma\left(\frac{j+1}{2}\right) \left[1 + (-1)^j \tilde{\gamma}\left(\frac{x^2}{2}, \frac{j+1}{2}\right)\right] \right\} \\ &= \sum_{j=0}^{N-1} \left\{ \frac{\exp\left(\frac{x^2}{2}\right)}{\Gamma(N-j)} \frac{\Gamma(N)}{\Gamma\left(\frac{N+1}{2}\right)} \frac{\Gamma\left(\frac{j+1}{2}\right)}{\Gamma(j+1)} \frac{x^{N-1-j}}{2^{\frac{N-j}{2}}} \left[1 + (-1)^j \tilde{\gamma}\left(\frac{x^2}{2}, \frac{j+1}{2}\right)\right] \right\} \\ &\overset{(a)}{=} \sum_{j=0}^{N-1} \left\{ \frac{\exp\left(\frac{x^2}{2}\right)}{\Gamma(N-j)} \frac{2^{N-1}\,\Gamma\left(\frac{N}{2}\right)}{\sqrt{\pi}} \frac{2^{-j}\sqrt{\pi}}{\Gamma\left(\frac{j}{2}+1\right)} \frac{x^{N-1-j}}{2^{\frac{N-j}{2}}} \left[1 + (-1)^j \tilde{\gamma}\left(\frac{x^2}{2}, \frac{j+1}{2}\right)\right] \right\} \\ &\overset{(b)}{=} \sum_{j=0}^{N-1} \exp\big(d(N, j, x)\big)\end{aligned}$$

where (a) follows from the equality

$$\Gamma(2u) = \frac{2^{2u-1}}{\sqrt{\pi}} \Gamma(u) \Gamma\left(u + \frac{1}{2}\right), \quad u \neq 0, -\frac{1}{2}, -1, -\frac{3}{2}, \ldots$$

and (b) follows from the definition of the function d in (5.74).

Chapter 6

Summary and Outlook

This research work focuses on the fundamental tradeoff between the performance of graph-based codes and the complexity required to achieve this performance under iterative message-passing decoding algorithms. The complexity of message-passing algorithms is dictated by two main factors – the complexity of a single decoding iteration and the number of iterations required to achieve the desired performance level. In this work, we derive lower bounds on both of these factors. These bounds refer to the asymptotic case where we let the block length of the codes tend to infinity. We also consider the fundamental performance limitations of finite-length block codes. This is done via the derivation of an improved sphere-packing (ISP) bound which applies to block codes transmitted over memoryless symmetric channels.

6.1 Contributions of this Dissertation

In Chapter 2 we derive two types of bounds. The first category consists of lower bounds on the asymptotic density of parity-check matrices of binary linear block codes, which are given in terms of the gap between the code rate and the channel capacity. Due to the nature of the iterative message-passing decoding algorithms, these bounds serve as lower bounds on the complexity (per information bit) of a single decoding iteration. The second category of bounds are upper bounds on the achievable rates of binary linear block codes (even under maximum-likelihood (ML) decoding). The bounds in both categories refer to the asymptotic case where we let the block length of the codes tend to infinity and assume that the transmission takes place over an arbitrary memoryless binary-input output-symmetric (MBIOS) channel. The derivation of the bounds was motivated by the desire to tighten the statements in [17, Theorems 1 and 2] and [81, Theorem 2.1]. The two-level quantization of the information on the log-likelihood ratio (LLR) in [17, 81] in essence performs the

analysis on a physically degraded binary symmetric channel (BSC) instead of the original communication channel. As a first step, we present an analysis based on information from a quantized channel which better reflects the statistics of the actual communication channel (though the quantized information is still degraded w.r.t. the original information provided by the channel). The number of quantization levels applied to the information is set as an arbitrary integer power of 2. The calculation of the bounds is subject to an optimization of the quantization levels of the LLR, as to get the tightest bounds within their form. Later, we present bounds that rely on the conditional pdf of the LLR at the output of the MBIOS channel, and perform the analysis on an equivalent channel without a degradation of the channel information. This second approach finally leads to bounds which are uniformly tighter than the bounds based on a quantization of the communication channel. It appears to be even simpler to calculate the un-quantized bounds, as their calculation does not involve the solution of any optimization equation related to the quantization levels. The comparison between the quantized and un-quantized bounds gives insight on the effect of the number of quantization levels of the LLR (even if they are chosen optimally) on the achievable rates, as compared to the ideal case where no quantization is done.

The information-theoretic bounds derived in Chapter 2 are valid for *every* sequence of binary linear block codes, in contrast to high probability results which follow from probabilistic tools (e.g., density evolution (DE) analysis under iterative message-passing decoding). The bounds hold under ML decoding, and hence, they hold in particular under any sub-optimal decoding algorithm. We apply the bounds to ensembles of low-density parity-check (LDPC) codes where the significance of these bounds is as follows: Firstly, by comparing the new upper bounds on the achievable rates with thresholds provided by DE analysis, we obtain rigorous bounds on the loss in performance of various LDPC ensembles due to the sub-optimality of message-passing decoding (as compared to ML decoding). Secondly, the parity-check density of binary linear block codes which are represented by standard bipartite graphs is interpreted as the complexity per iteration under message-passing decoding. Therefore, by tightening the reported lower bound on the asymptotic parity-check density (see [81, Theorem 2.1]), the new bounds provide better insight on the tradeoff between the asymptotic performance and the asymptotic decoding complexity of iteratively decoded LDPC codes. Thirdly, the new lower bound on the bit error probability of binary linear block codes (see Corollary 2.4) tightens the reported lower bound in [81, Theorem 2.5] and provides a quantitative measure to the number of fundamental cycles in the graph which should exist in terms of the achievable rate (even under ML decoding) and its gap to capacity. It is well known that cycle-free codes have poor

performance [103], so the lower bound on the minimal number of fundamental cycles in the graph (i.e., cycles which cannot be decomposed into some more elementary cycles) as a function of the gap in rate to capacity strengthens the result in [103] on the inherent limitation of cycle-free codes.

The lower bound on the asymptotic parity-check density in [81, Theorem 2.1] and its improvements in Chapter 2 grow like the log of the inverse of the gap (in rate) to capacity. The result in [81, Theorem 2.2] shows that a logarithmic growth rate of the parity-check density is achievable for Gallager's regular LDPC ensemble under ML decoding when transmission takes place over an arbitrary MBIOS channel. These results show that for any iterative decoder which is based on the representation of the codes by Tanner graphs, there exists a tradeoff between asymptotic performance and complexity which cannot be surpassed. Recently, it was shown in [64, 65] that a better tradeoff can be achieved by allowing more complicated graphical models which involve a sufficient number of state nodes in the graph; for the particular case of the binary erasure channel (BEC), the encoding and the decoding complexity of properly designed codes on graphs remains bounded as the gap to capacity vanishes.

The analysis in Chapter 3 generalizes the statements in Chapter 2 to the case where the codes are transmitted over a set of statistically independent parallel MBIOS channels. The bounds on the asymptotic achievable rates and parity-check density can be applied to various scenarios which form particular cases of communication over parallel channels, e.g., intentionally punctured LDPC codes [35], non-uniformly error protected LDPC codes [69], and LDPC-coded modulation (see e.g., [37, 112]). In Section 3.4, we use Theorem 3.1 for the derivation of upper bounds on the achievable rates under ML decoding of (randomly and intentionally) punctured LDPC codes whose transmission takes place over an MBIOS channel. It is exemplified numerically that for various good ensembles of intentionally punctured LDPC codes, the asymptotic loss in performance due to the code structure is still non-negligible as compared to the corresponding loss due to the sub-optimality of iterative decoding (as compared to optimal ML decoding). Looser versions of the bounds derived in this chapter for punctured LDPC codes suggest a simplified re-derivation of previously reported bounds on the decoding complexity of randomly punctured LDPC codes (see [65, Theorems 3 and 4]).

In Chapter 4, we turn to consider the number of iterations required for successful decoding of graph-based code ensembles. In the considered setting, we let the block length of these ensembles tend to infinity, and the transmission takes place over a BEC. Theorems 4.1–4.3 demonstrate that for various attractive families of code

ensembles (including LDPC codes, systematic and non-systematic irregular repeat-accumulate (IRA) codes, and accumulate-repeat-accumulate (ARA) codes), the number of iterations which is required to achieve a desired bit erasure probability scales at least like the inverse of the gap between the channel capacity and the design rate of the ensemble. This conclusion holds provided that the fraction of degree-2 variable nodes in the Tanner graph does not tend to zero as the gap to capacity vanishes (where under mild conditions, this property is satisfied for sequences of capacity-achieving LDPC ensembles, see [76, Lemma 5]). The behavior of these lower bounds matches well with the experimental results and the conjectures on the number of iterations and complexity, as provided by Khandekar and McEliece (see [42, 43, 56]). In [42, Theorem 3.5], it was stated that for LDPC and IRA ensembles which achieve a fraction $1-\varepsilon$ of the channel capacity of a BEC with a target bit erasure probability of P_{b} under iterative message-passing decoding, the number of iterations grows like $O\left(\frac{1}{\varepsilon}\right)$. In light of the outline of the proof of this statement, as suggested in [42, p. 71], it implicitly assumes that the flatness condition is satisfied for these code ensembles and also that the target bit erasure probability vanishes; under these assumptions, the reasoning suggested by Khandekar in [42, Section 3.6] supports the behavior of the lower bounds which are derived in this work.

The lower bounds on the number of iterations in Chapter 4 become trivial when the fraction of degree-2 variable nodes vanishes. However, this is mainly a result of our focus on the derivation of simple lower bounds on the number of iterations which do not depend on the full characterization of the degree distributions of the code ensembles. Following the proofs of the statements in this chapter and focusing on the case where the fraction of degree-2 variable nodes vanishes, it is possible to derive lower bounds on the number of iterations which are not trivial even in this case; these bounds, however, require the knowledge of the entire degree distribution of the examined ensembles.

To complement the asymptotic analysis in Chapters 2–4, we study the fundamental limitations on the performance of finite-length block codes. Chapter 5 presents an improved sphere-packing (ISP) bound for finite-length block codes whose transmission takes place over symmetric memoryless channels. The derivation of the ISP bound was stimulated by the remarkable performance and feasible complexity of turbo-like codes with short to moderate block lengths. We were motivated by recent improvements on the 1967 sphere-packing (SP67) bound (see [87]) for finite block lengths, as suggested by Valembois and Fossorier [109]. Numerical results demonstrate that the ISP bound significantly tightens the SP67 bound and the bound in [109, Theorem 7]. This improvement is especially pronounced for codes of short to moderate

block lengths and also grows as the input alphabet of the considered communication channel increases in size.

The ISP bound is applied to M-ary phase shift keying (PSK) block coded modulation schemes whose transmission takes place over an additive white Gaussian noise (AWGN) channel and the received signals are coherently detected. The tightness of the ISP bound is exemplified by comparing it with upper and lower bounds on the ML decoding error probability and also with reported computer simulations of turbo-like codes under iterative decoding. Chapter 5 also presents a new algorithm which performs the entire calculation of Shannon's 1959 sphere-packing (SP59) bound (see [89]) in the logarithmic domain, thus facilitating the exact calculation of the SP59 bound for all block lengths without the need for asymptotic approximations. It is shown that the ISP bound suggests an interesting alternative to the SP59 bound, where the latter is specialized for the AWGN channel.

6.2 Future Research Directions

Performance versus complexity of generalized LDPC codes

The component codes of LDPC codes can be viewed as repetition codes at the variable nodes and single-parity check (SPC) codes at the check nodes. The idea in the construction of generalized LDPC (GLDPC) codes is to replace a fraction of the SPC codes by other algebraic codes (e.g., Hamming or BCH codes). GLDPC codes were introduced independently by Lentmaier and Zigangirov [46, 47], Boutros et al. [16] and Wadayama [111]. GLDPC codes are constructed by replacing each single parity-check in Gallager's LDPC codes with the parity-check matrix of a small linear block code called a constituent code. It has been shown that GLDPC codes are asymptotically good in the sense of their minimum distance, and even more importantly, they exhibit with iterative decoding algorithms good performance over both AWGN channel and Rayleigh fading channels. Moreover, it was demonstrated that GLDPC codes can also be considered as a generalization of product codes, and because of their flexibility on the selection of code length, GLDPC codes turn out to be a promising alternative to product codes in many applications.

The design of doubly generalized low-density parity-check (DGLDPC) codes with generic block linear codes at both bit and check nodes (instead of the traditional repetition and single parity-check codes) was recently considered in [114]. For the binary BEC and the AWGN channel. Both analysis and simulations show that this approach provides more flexibility in constructing codes with good threshold. An

improvement in their performance was exemplified at the expense of increasing their encoding and decoding complexity [114]. It would be interesting to compare the performance of LDPC and generalized LDPC cods under iterative decoding when their decoding complexity is fixed, and to devise some approaches for the optimization of generalized LDPC codes when one wishes to find a good tradeoff between the asymptotic threshold and the decoding complexity under iterative decoding.

Application of the information-theoretic bounds for binary linear block codes over parallel channels to common communication scenarios

In Chapter 3, we derive information-theoretic bounds on the performance-complexity tradeoff for LDPC ensembles whose transmission takes place over a set of parallel MBIOS channels. In light of the widespread use of punctured LDPC codes, these bounds are to assess the performance and complexity of randomly and intentionally punctured LDPC ensembles transmitted over a single MBIOS channel. Parallel channels serve as a model for several other important communication scenarios; these include non-uniformly error-protected codes, transmission over block-fading channels and multi-carrier signaling. It is suggested to apply the bounds derived in Chapter 3 towards providing information-theoretic bounds on the performance and complexity of LDPC ensembles for these applications. It would also be interesting to compare the theoretic bounds to performance of existing schemes for these scenarios.

Generalization of the lower bounds on the number of decoding iterations to arbitrary MBIOS channels

In Chapter 4, we derive lower bounds on the number of iterations which is required for successful decoding of several families of graph-based code ensembles. These bounds refer to codes transmitted over a BEC, and they hold in the asymptotic case where the block length of the codes tends to infinity. The derivation of the bounds relies on extrinsic information-transfer (EXIT) functions and the area theorem, and hinges on the fact that EXIT charts coincide with density-evolution analysis for the BEC. As a topic for further research, it is suggested to examine the possibility of adapting the bounds presented in Chapter 4 to the case where the transmission takes place over arbitrary MBIOS channels. This direction is especially appealing in light of the recent development of generalized EXIT charts and the introduction of the generalized area theorem [59], which apply to arbitrary MBIOS channels.

Further improvement of sphere-packing bounds for finite-length codes

In Chapter 5, we present a new sphere-packing lower bound on the error probability of optimal block codes. This improved sphere-packing (ISP) bound is targeted at finite-length block codes transmitted over symmetric memoryless channels. It is demonstrated that the ISP bound significantly tightens the classical 1967 sphere-packing (SP67) bound and its recent improved version in [109]; this is especially pronounced for codes for short to moderate block lengths. However, it is also demonstrated that for M-ary PSK modulated signals transmitted over the AWGN channel, the classical 1959 sphere-packing (SP59) bound of Shannon may provide a tighter lower bound, especially for short codes with low rates. This is despite the fact that the SP59 does not take into account the specific modulation used and only assumes that the signals have equal energy. This observation motivates further improvements of the technique, in order to further tighten the bound for finite-length codes. Such an improvement might be achieved by optimizing the tilting measure f_s (see Eq. 5.50) for the specific code length considered, instead of using the function from [87], which is optimized for the asymptotic case where the block length tends to infinity. It would also be interesting to revisit the derivation of the bound and consider the case where the communication channel is not symmetric.

Sphere-packing bounds on the symbol error probability of optimal codes

The sphere-packing bounds in [87, 89, 109] and Chapter 5 refer to the *word* error probability of optimal block codes. In many applications, the relevant distortion measure is the bit (or symbol) error probability. Moreover, many graph-based codes exhibit very low bit error probabilities under iterative decoding, in contrast to a rather high probability of word error. As a topic for further research, it is suggested to examine the possibility of adapting the sphere-packing bounding techniques to derive lower bounds on the symbol error probability of optimal codes.

References

[1] Big-O notation. [Online]. Available: http://www.nist.gov/dads/HTML/bigOnotation.html.

[2] IEEE standard for information technology-telecommunications and information exchange between systems-local and metropolitan area networks-specific requirements - part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications.

[3] IEEE standard for local and metropolitan area networks part 16: Air interface for fixed broadband wireless access systems," *IEEE Std 802.16-2004 (Revision of IEEE Std 802.16-2001)*.

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

[5] B. Ammar, Y. Kou, J. Xu, and S. Lin, "Construction of low-density parity-check codes based on balanced incomplete block designs," *IEEE Transactions on Information Theory*, vol. 50, no. 6, pp. 1257–1268, June 2004.

[6] A. Amraoui, A. Montanari, and R. Urbanke, "How to find good finite-length codes: from art towards science," *European Transactions on Telecommunications*, vol. 18, no. 5, pp. 491–508, August 2007.

[7] K. Andrews, V. Stanton, S. Dolinar, V. Chen, J. Berner, and F. Pollara, "Turbo-decoder implementation for the deep space network," *Interplanetary Network Progress Report*, IPN Progress Report 42-148, October–December 2001. [Online]. Available: http://tmo.jpl.nasa.gov/progress_report/42-148/148A.pdf.

[8] M. Ardakani, B. Smith, W. Yu, and F. R. Kschischang, "Complexity-optimized low-density parity-check codes," in *Proceedings of the Forty-Third Annual Allerton Conference on Communication, Control and Computing*, pp. 45–54, Urbana-Champaign, Illinois, USA, September 28–30 2005.

[9] A. Ashikhmin, G. Kramer, and S. ten Brink, "Extrinsic information transfer functions: Model and erasure channel properties," *IEEE Transactions on Information Theory*, vol. 50, no. 11, pp. 2657–2673, November 2004.

[10] O. Barak, D. Burshtein, and M. Feder, "Bounds on achievable rates of LDPC codes used over the binary erasure channel," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2483–2489, October 2004.

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

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

[13] E. R. Berlekamp, "The performance of block codes," Notices of the AMS, pp. 17–22, January 2002. [Online]. Available: http://www.ams.org/notices/200201/fea-berlekamp.pdf.

[14] C. Berrou, "The ten-year-old turbo codes are entering into service," *IEEE Communications Magazine*, vol. 41, no. 8, pp. 110–116, August 2003.

[15] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near shannon limit error-correcting coding and decoding: Turbo codes," in *Proceedings of the 1993 IEEE International Conference on Communication (ICC 93)*, Geneva, Switzerland, May 23–26, 1993, pp. 1064–1070.

[16] J. Boutros, O. Pothier, and G. Zemor, "Generalized low-density (Tanner) codes," in *Proceedings of the 1999 IEEE International Conference on Communication (ICC 1999)*, pp. 441–445, Vancouver, British Columbia, Canada, June 6–10, 1999.

[17] D. Burshtein, M. Krivelevich, S. Litsyn, and G. Miller, "Upper bounds on the rate of LDPC codes," *IEEE Transactions on Information Theory*, vol. 48, no. 9, pp. 2437–2449, September 2002.

[18] S.-Y. Chung, T. J. Richardson, and R. Urbanke, "Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 657–670, February 2001.

[19] S.-Y. Chung, G. D. Forney Jr., T. J. Richardson, and R. Urbanke, "On the design of low-density parity-check codes within 0.0045 db of the Shannon limit," *IEEE Communication Letters*, vol. 5, no. 2, pp. 58–60, February 2001.

[20] D. J. Costello and G. D. Forney Jr., "Channel coding: The road to channel capacity," *Proceedings of the IEEE*, vol. 95, no. 6, pp. 1150–1177, June 2007.

[21] J. Cuevas, P. Adde, and S. Kerouedan, "Turbo decoding of product codes for gigabit per second applications and beyond," *European Transactions on Telecommunications*, vol. 17, no. 1, pp. 45–55, Jan.–Feb. 2006.

[22] C. Di, D. Proietti, I. E. Telatar, and R. Urbanke, "Finite-length analysis of low-density parity-check codes on the binary erasure channel," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1570–1579, June 2002.

[23] D. Divsalar and S. Doliner, "Concatenation of hamming codes and accumulator codes with high-order modulations for high-speed decoding," Jet Propulsion Laboratory (JPL), IPN Progress Report 42-156, February 2004. [Online]. Available: http://tmo.jpl.nasa.gov/progress_report/42-156/156G.pdf.

[24] D. Divsalar, H. Jin, and R. J. McEliece, "Coding theorems for 'turbo-like' codes," in *Proceedings of the Thirty-Sixth Annual Allerton Conference on Communication, Control and Computing*, pp. 201–210, Urbana-Champaign, IL, USA, September 23-25, 1998.

[25] D. Divsalar and C. Jones, "Protograph LDPC codes with node degrees at least 3," in *Proceedings of the 2006 IEEE Global Communications Conference (GlobeCom 2006)*, pp. 1–5, San Francisco, CA, USA, November 27–December 1 2006.

[26] D. Divsalar, C. Jones, S. Doliner, and J. Thorpe, "Protograph based LDPC codes with minimum distance linearly growing with block size," in *Proceedings of the 2005 IEEE Global Communications Conference (GlobeCom 2005)*, pp. 1152–1156, St. Louis, MO, USA, November 28–December 2 2005.

[27] S. Doliner, D. Divsalar, and F. Pollara, "Code performance as a function of block size," Jet Propulsion Laboratory (JPL), TMO Progress Report 42-133, May 1998. [Online]. Available: http://ipnpr.jpl.nasa.gov/progress_report/42-133/133K.pdf.

[28] P. Elias, "Error-free coding," *IRE Transactions on Information Theory*, vol. 4, pp. 29–37, September 1954.

[29] ——, "List decoding for noisy channels," Research Laboratory of Electronics, Massachusetts Institute of Technology (MIT), Tech. Rep. 335, September 1957.

[30] R. G. Gallager, "Low-density parity-check codes," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA, 1963.

[31] ——, "A simple derivation of the coding theorem and some applications," *IEEE Transactions on Information Theory*, vol. 11, no. 1, pp. 3–18, January 1965.

[32] ——, *Information Theory and Reliable Communications*. John Wiley and Sons, 1968.

[33] V. Guruswami, *Algorithmic Results in List Decoding*, Foundations and Trends in Theoretical Computer Science. Now Publishers, vol. 2, no. 2, pp. 107–195, December 2006.

[34] J. Ha, "Low-density parity-check codes with erasures and puncturing," Ph.D. dissertation, Georgia Institute of Technology, Atlanta, GA, USA, November 2003.

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

[36] H. Herzberg and G. Poltyrev, "The error probability of m-ary psk block coded modulation schemes," *IEEE Transactions on Communications*, vol. 44, no. 4, pp. 427–433, April 1996.

[37] J. Hou, P. H. Siegel, L. B. Milstein, and H. D. Pfister, "Capacity-approaching bandwidth-efficient coded modulation schemes based on low-density parity-check codes," *IEEE Transactions on Information Theory*, vol. 49, no. 9, pp. 2141–2155, September 2003.

[38] C. H. Hsu and A. Anastasopoulos, "Capacity-achieving codes for noisy channels with bounded graphical complexity and maximum-likelihood decoding," submitted to *IEEE Transactions on Information Theory*, March 2006. [Online]. Available: http://www.eecs.umich.edu/~anastas/docs/tit06_ha.pdf.

[39] ——, "Capacity-achieving LDPC codes through puncturing," submitted to *IEEE Transactions on Information Theory*, December 2006. [Online]. Available: http://www.eecs.umich.edu/~anastas/docs/tit06b_ha.pdf.

[40] H. Jin, A. Khandekar, and R. J. McEliece, "Irregular repeat-accumulate codes," in *Proceedings of the Second International Symposium on Turbo Codes & Related Topics*,pp. 1–8, Brest, France, September 4-7 2000.

[41] H. Jin and R. J. McEliece, "Typical pairs decoding on the AWGN channel," in *Proceedings of the 2000 International Symposium on Information Theory and Its Applications*, pp. 180–183, Honolulu, Hawaii, USA, November 5–8, 2000.

[42] A. Khandekar, "Graph-based codes and iterative decoding," Ph.D. dissertation, California Institute of Technology, Pasadena, CA, USA, June 2002. [Online]. Available: http://etd.caltech.edu/etd/available/etd-06202002-170522.

[43] A. Khandekar and R. J. McEliece, "On the complexity of reliable communications on the erasure channel," in *Proceedings of the 2001 IEEE International Symposium on Information Theory (ISIT 2001)*, p. 1, Washington, District of Columbia, USA, June 24–29, 2001.

[44] D. E. 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*, pp. 192–195, Brest, France, September 3–5 1997.

[45] D. E. Lazic, T. Beth, and S. Egner, "Constrained capacity of the AWGN channel," in *Proceedings of the 1998 IEEE International Symposium on Information Theory (ISIT 1998)*, p. 237, Cambridge, MA, USA, August 16–21 1998.

[46] M. Lentmaier, "Soft iterative decoding of generalized low-density parity-check codes based on map decoding of component hamming codes," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA, 1963.

[47] M. Lentmaier and K. S. Zigangirov, "On generalized low-density parity-check codes based on hamming component codes," *IEEE Communication Letters*, vol. 3, no. 8, pp. 248–250, August 1999.

[48] S. Lin and D. J. Costello, *Error Control Coding*, 2nd edition, Prentice Hall, 2004.

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

[50] Y. Liu, J. Hou, and V. K. N. Lau, "Complexity bounds of LDPC codes for parallel channels," in *Proceedings of the Forty-Second Annual Allerton Conference on Communication, Control and Computing*, pp. 1705–1713, Urbana-Champaign, IL, USA, September 29–October 1, 2004.

[51] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. A. Spielman, "Efficient erasure correcting codes," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 569–584, February 2001.

[52] X. Ma and E. Yang, "Low-density parity-check codes with fast decoding convergence speed," in *Proceedings of the 2004 IEEE International Symposium on Information Theory (ISIT 2004)*, Chicago, Illinois, USA, June 27-July 2 2004, p. 277, a preprint of a full paper version of this work is available at http://www.arxiv.org/abs/cs.IT/0602081.

[53] D. J. C. MacKay, "A free energy minimization framework for inference problems in modulo 2 arithmetic," in *Fast Software Encryption (Proceedings of the of the 1994 K.U. Leuven Workshop on Cryptographic Algorithms)*, Lectures Notes in Computer Science, B. Preneel, Editor, vol. 1008, pp. 179–195 Springer Verlag, 1995.

[54] D. J. C. MacKay and R. M. Neal, "Good codes based on very sparse matrices," *IEEE Transactions on Information Theory*, vol. 45, no. 2, pp. 399–431, March 1999.

[55] S. J. Macmullan and O. M. Collins, "A comparison of known codes, random codes and the best codes," *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 3009–3022, November 1998.

[56] R. J. McEliece, "Are turbo-like codes effective on non-standard channels?" *IEEE Information Theory Society Newsletter*, vol. 51, no. 4, pp. 1–8, December 2001.

[57] C. Measson, "Conservation laws for coding," Ph.D. dissertation, EPFL, Lausanne, Switzerland, March 2006. [Online]. Available: http://lthcwww.epfl.ch/~cyril/research/These_3485_Measson_Book.pdf.

[58] C. Measson, A. Montanari, T. Richardson, and R. Urbanke, "Life above threshold: From list decoding to area theorem and MSE," in *Proceedings of the 2004 IEEE Information Theory Workshop*, San Antonio, Texas, USA, October 24–29, 2004.

[59] ——, "The generalized area theorem and some of its consequences," submitted to *IEEE Transactions on Information Theory*, November 2005. [Online]. Available: http://arxiv.org/abs/cs/0511039.

[60] C. Measson, A. Montanari, and R. Urbanke, "Maxwell construction: The hidden bridge between iterative and maximum a posteriori decoding," accepted to *IEEE Transactions on Information Theory*. [Online]. Available: http://www.arxiv.org/abs/cs.IT/0506083.

[61] A. Montanari, personal communications, May 2005.

[62] ——, "Tight bounds for LDPC codes and LDGM codes under MAP decoding," *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3221–3246, September 2005.

[63] P. Oswald and A. Shokrollahi, "Capacity-achieving sequences for the erasure channel," *IEEE Transactions on Information Theory*, vol. 48, no. 12, pp. 3017–3028, December 2002.

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

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

[66] L. Ping, X. Huang, and N. Phamdo, "Zigzag codes and concatenated zigzag codes," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 800–807, February 2001.

[67] H. Pishro-Nik and F. Fekri, "On decoding of low-density parity-check codes over the binary erasure channel," *IEEE Transactions on Information Theory*, vol. 50, no. 3, pp. 439–454, March 2004.

[68] ——, "Results on puctured low-density parity-check codes and improved iterative deocding techniques," *IEEE Transactions on Information Theory*, vol. 53, no. 2, pp. 599–614, February 2007.

[69] H. Pishro-Nik, N. Rahnavard, and F. Fekri, "Nonuniform error correction using low-density parity-check codes," *IEEE Transactions on Information Theory*, vol. 51, no. 7, pp. 2702–2714, July 2005.

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

[71] T. Richardson, A. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 619–637, February 2001.

[72] T. Richardson and R. Urbanke, "Finite-length density evolution and the distribution of the number of iterations for the binary erasure channel," unpublished.

[73] ——, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 599–618, February 2001.

[74] ——, *Modern Coding Theory*. Cambridge University Press, 2008, to be published. [Online]. Available: http://lthcwww.epfl.ch/mct/index.php.

[75] R. M. Roth, *Introduction to Coding Theory*. Cambridge University Press, 2006.

[76] I. Sason, "On universal properties of capacity-approaching LDPC ensembles," submitted to *IEEE Transactions on Information Theory*, September 2007. [Online]. Available: http://www.arxiv.org/abs/0709.0599.

[77] I. Sason and I. Goldenberg, "Coding for parallel channels: Gallager bounds and applications to turbo-like codes," *IEEE Transactions on Information Theory*, vol. 53, no. 7, pp. 2394–2428, July 2007.

[78] 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 Transactions on Information Theory*, vol. 47, no. 6, pp. 2275–2299, September 2001.

[79] ——, *Performance Analysis of Linear Codes under Maximum-Likelihood Decoding: A Tutorial*, Foundations and Trends in Communication and Information Theory, Now Publishers, vol. 3, no. 1–2, pp. 1–222, June 2006. [Online]. Available: http://www.ee.technion.ac.il/people/sason/monograph_postprint.pdf.

[80] I. Sason and R. Urbanke, "Complexity versus performance of capacity-achieving irregular repeat-accumulate codes on the binary erasure channel," *IEEE Transactions on Information Theory*, vol. 50, no. 6, pp. 3017–3028, December 2002.

[81] ——, "Parity-check density versus performance of binary linear block codes over memoryless symmetric channels," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1611–1635, July 2003.

[82] I. Sason and G. Wiechman, "Log-domain calculation of the 1959 sphere-packing bound with application to M-ary PSK block coded modulation,"

in *Proceedings of the 24th IEEE Convention of Electrical and Electronics Engineers in Israel*, pp. 344 – 348, Eilat, Israel, November 15 - 17 2006.

[83] ——, "On achievable rates and complexity of LDPC codes for parallel channels: Information-theoretic bounds and applications," in *Proceedings of the 2006 IEEE International Symposium on Information Theory (ISIT 2006)*, pp. 406–410, Seattle, Washington, USA, July 9-14 2006. [Online]. Available: http://arxiv.org/abs/cs.IT/0512076.

[84] ——, "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. [Online]. Available: http://arxiv.org/abs/cs.IT/0512075

[85] ——, "On achievable rates and complexity of LDPC codes over parallel channels: Bounds and applications," *IEEE Transactions on Information Theory*, vol. 53, no. 2, pp. 580–598, February 2007.

[86] ——, "Bounds on the number of iterations for turbo-like ensembles over the binary erasure channel," submitted to *IEEE Transactions on Information Theory*, November 2007. [Online]. Available: http://www.arxiv.org/abs/0711.1056.

[87] C. Shannon, R. Gallager, and E. Berlekamp, "Lower bounds to error probability for decoding on discrete memoryless channels," *Information and Control*, vol. 10, pp. 65–103 (Part 1), and 522–552 (Part 2), February / May 1967.

[88] C. E. Shannon, "A mathematical theory of communications," *Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, July / October 1948.

[89] ——, "Probability of error for optimal codes in a Gaussian channel," *Bell System Technical Journal*, vol. 38, no. 3, pp. 611–656, May 1959.

[90] R. Y. Shao, S. Lin, and M. P. C. Fossorier, "Two simple stopping criteria for turbo decoding," *IEEE Transactions on Communications*, vol. 47, no. 8, pp. 1117–1120, August 1999.

[91] E. Sharon, A. Ashikhmin, and S. Litsyn, "EXIT functions for binary input memoryless symmetric channels," *IEEE Transactions on Communications*, vol. 54, no. 7, pp. 1207–1214, July 2006.

[92] E. Sharon, S. Litsyn, and J. Goldenberg, "Efficient serial message-passing schedules for LDPC decoding," *IEEE Transactions on Information Theory*, vol. 53, no. 11, pp. 4076–4091, November 2007.

[93] A. Shokrollahi, "New sequences of time erasure codes approaching channel capacity," in *Proceedings of the of the 13th International Symposium on Applied Algebra, Algebraic Algorithms and Error-Correcting Codes*, Lectures Notes in Computer Science, vol. 1719, pp. 65–76. Springer Verlag, 1999.

[94] ——, "Capacity-achieving sequences," *IMA Volume in Mathematics and its Applications*, vol. 123, pp. 153–166, 2000.

[95] ——, "Raptor codes," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2551–2567, June 2006.

[96] N. J. A. Sloane and A. D. Wyner, Eds., *Claude Elwood Shannon – Collected Papers*. IEEE Press, 1993.

[97] Y. Tai, L. Lan, L. Zeng, S. Lin, and K. Abdel-Ghaffar, "Algebraic construction of quasi-cyclic LDPC codes for the AWGN and erasure channels," *IEEE Transactions on Communications*, vol. 54, no. 10, pp. 1756–1765, October 2006.

[98] O. Y. Takeshita, O. M. Collins, P. C. Massey, and D. J. Costello, "On the frame-error rate of concatenated turbo codes," *IEEE Transactions on Communications*, vol. 49, no. 4, pp. 602–608, April 2001.

[99] H. Tang, J. Xu, S. Lin, and K. Abdel-Ghaffar, "Codes on finite geometries," *IEEE Transactions on Information Theory*, vol. 51, no. 2, pp. 572–596, February 2005.

[100] R. M. Tanner, "A recursive approach to low-complexity codes," *IEEE Transactions on Information Theory*, vol. 27, no. 5, pp. 533–547, September 1981.

[101] S. ten Brink, "Code characteristic matching for iterative decoding of serially concatenated codes," *Annals of Telecommunications*, vol. 56, no. 7-8, pp. 394–408, July-August 2001.

[102] ——, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Transactions on Communications*, vol. 49, no. 10, pp. 1727–1737, October 2001.

[103] A. Trachtenberg, T. Etzion, and A. Vardy, "Which codes have cycle-free Tanner graphs?" *IEEE Transactions on Information Theory*, vol. 45, no. 6, pp. 2173–2181, September 1999.

[104] M. Twitto and I. Sason, "On the error exponents of some improved tangential-sphere bounds," *IEEE Transactions on Information Theory*, vol. 53, no. 3, pp. 1196–1210, March 2007.

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

[106] R. Urbanke. Error floor calculator for the binary erasure channel. [Online]. Available: http://lthcwww.epfl.ch/research/efc.

[107] ——. Optimization of degree distributions for ensembles of LDPC codes. [Online]. Available: http://lthcwww.epfl.ch/research/ldpcopt/index.php.

[108] A. Valembois and M. Fossorier, "Box and match techniques applied to soft-decision decoding," *IEEE Transactions on Information Theory*, vol. 50, no. 5, pp. 796–810, May 2004.

[109] ——, "Sphere-packing bounds revisited for moderate block length," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 2998–3014, December 2004.

[110] A. J. Viterbi and J. K. Omura, *Principles of Digital Communications and Coding*. McGraw-Hill Book Company, 1979.

[111] T. Wadayama, "An extension of Gallager ensemble of low-density parity-check codes," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E85–A, no. 1, pp. 1161–1171, January 2002.

[112] ——, "Ensemble analysis on syndrome entropy of binary linear codes," in *Proceedings of the 2006 IEEE International Symposium on Information Theory (ISIT 2006)*, pp. 1559–1563, Seattle, Washington, USA, July 9–14 2006.

[113] C. C. Wang, S. R. Kulkarni, and H. V. Poor, "Finite-dimensional bounds on $\mathbb{Z}_m$ and binary LDPC codes with belief-propagation decoders," *IEEE*

Transactions on Information Theory, vol. 53, no. 1, pp. 56–81, January 2007.

[114] Y. Wang and M. Fossorier, "Doubly generalized LDPC codes," in *Proceedings of the 2006 IEEE International Symposium on Information Theory (ISIT 2006)*, pp. 669–673, Seattle, Washington, USA, July 9–14 2006.

[115] L. Wei, "Near-optimum serial concatenation of single-parity codes with convolutional codes," *IEE Proceedings on Communications*, vol. 152, no. 4, pp. 397–403, August 2005.

[116] G. Wiechman and I. Sason, "On the parity-check density and achievable rates of LDPC codes for memoryless binary-input output-symmetric channels," in *Proceedings of the Forty-Third Annual Allerton Conference on Communication, Control and Computing*, pp. 1747–1758, Urbana-Champaign, IL, USA, September 28-30 2005. [Online]. Available: http://arxiv.org/abs/cs.IT/0505078

[117] ——, "An improved sphere-packing bound targeting codes of short to moderate block lengths and applications," in *Proceedings of the Forty-Third Annual Allerton Conference on Communication, Control and Computing*, pp. 1–12, Urbana-Champaign, IL, USA, September 27-29 2006.

[118] ——, "Parity-check density versus performance of binary linear block codes: New bounds and applications," *IEEE Transactions on Information Theory*, vol. 53, no. 2, pp. 550–579, February 2007.

[119] ——, "An improved sphere-packing bound for finite-length error-correcting codes," submitted to *IEEE Transactions on Information Theory*, March 2007. [Online]. Available: http://www.arxiv.org/abs/cs.IT/0608042.

[120] J. M. Wozencraft, "List decoding," Research Laboratory of Electronics, Massachusetts Institute of Technology (MIT)," Quarterly Progress Report, January 1958.

אספקטים תאורתיים ומעשיים הקשורים בברירה שבין ביצועים וסיבוכיות של קודי LDPC ונגזרותיהם

גיל וייכמן

אספקטים תאורתיים ומעשיים הקשורים בברירה שבין ביצועים וסיבוכיות של קודי LDPC ונגזרותיהם

חיבור על מחקר

לשם מילוי חלקי של הדרישות לקבלת תואר

דוקטור לפילוסופיה

גיל וייכמן

הוגש לסנט הטכניון — מכון טכנולוגי לישראל

שבט תשס"ח חיפה ינואר 2008

חיבור על מחקר נעשה בהדרכת ד"ר יגאל ששון
בפקולטה להנדסת חשמל

הכרת תודה

ברצוני להודות מקרב לב לדר' יגאל ששון על ההנחיה המסורה, ועל תרומתו הגדולה למחקר ולאנליזה המוצגים בעבודה זו. הוא תרם רבות להנאה ולחדוות היצירה שאפיינו את שנות השתלמותי בטכניון. היתה לי זכות גדולה לעבוד בשיתוף פעולה פורה ומהנה עם חוקר ברמתו. ברצוני להודות לו על שהחדיר בי מעט מסקרנותו ומרצונו העז להבין באופן מלא כל בעיה בה בחרנו לעסוק. מעבר לכך, היחס החם והחברי, סבלנותו ורצונו הכן לעזור ולתרום בכל בעיה מקצועית ואישית ילוו אותי לאורך שנים רבות.

ברצוני להודות להוריי, רבקה ומנחם על תמיכתם ועידודם. משחר ילדותי, הוצבו החינוך וההשכלה בראש סדר העדיפויות. עבודת דוקטורט זו הינה ללא ספק תוצאה ישירה של דגשים אלו.

עבודה זאת לא היתה מתאפשרת ללא תמיכתה החמה של אשתי, ענת. תודתי העמוקה נתונה לה על התמיכה והעידוד ועל ההבנה שגילתה במהלך ערבים ולילות רבים בהם הייתי נוכח-נפקד בביתנו. היא שימשה בת-זוג מושלמת בדרך הארוכה והמאתגרת בה הלכנו בשנים האחרונות.

אני מודה לארנה ואנדרו ויטרבי, לקרן נאמן ולטכניון על התמיכה הכספית הנדיבה בהשתלמותי.

המחקר נתמך על ידי הקרן הלאומית למדע (תקציב מס. 1070/07).

תקציר

קודים לתיקון שגיאות המשתמשים באלגוריתמי פענוח איטרטיביים נפוצים כיום כטכני-קות קידוד ערוץ בסיבוכיות נמוכה. קודים אלו ניתנים לתיאור על ידי מודלים גרפיים המשמשים הן לציון המבנה האלגברי של הקודים והן לאפיון פעולתם של אלגוריתמי הפע-נוח האיטרטיביים. בשנים האחרונות נתגלו משפחות רבות של קודים המאפשרים פענוח איטרטיבי יעיל. משפחות אלו כוללות קודים בעלי מטריצות בדיקת זוגיות דלילה (LDPC), קודים בעלי מטריצה יוצרת דלילה (LDGM), קודי טורבו, קודי מכפלה, קודי חזרה-אגירה (repeat-accumulate) ווריאנטים שלהם ועוד. קודים אלו מציגים תחת אלגוריתמי פענוח איטרטיביים ביצועים המתקרבים למגבלות הקיבול של ערוצי תקשורת סטנדרטיים רבים, בעודם שומרים על סיבוכיות פענוח מעשית. הביצועים יוצאי הדופן של קודים אלו מהווים מוטיבציה חזקה לחקירת המגבלות האינפורמציוניות של הברירה בין הביצועים לסיבוכיות הפענוח של קודים המבוססים על מודלים גרפיים. הביצועים של משפחות קודים אלו עבור אורכי בלוק מעשיים מהווים אף הם מוטיבציה לחקירת המגבלות האינפורמציוניות על ביצועיהם של קודי בלוק באורך סופי. המחקר בעבודה זו מונע על ידי שלוש שאלות עיקריות:

1. עד כמה טובים יכולים להיות הביצועים של קודי LDPC, אפילו כאשר משתמשים במפענח אופטימלי?
2. מהן המגבלות היסודיות על סיבוכיות הפענוח המינימלית של קודי גרף, כפונקציה של הפער בין קיבול הערוץ ובין קצב הקוד עבורו מושגת הסתברות שגיאה ששואפת לאפס?
3. מהן המגבלות היסודיות על ביצועיהם של קודי בלוק באורך סופי?

אלגוריתמי הפענוח האיטרטיביים פועלים על ידי שליחת הודעות לאורך קשתות הגרף המתאר את הקוד. משום כך, הסיבוכיות של אלגוריתמים אלו מושפעת משני גורמים עיקריים: הגורם הראשון הינו הסיבוכיות של המודל הגרפי המתאר את הקוד. גודל זה, המוגדר כמספר הקשתות בגרף מנורמל במספר סיביות האינפורמציה של הקוד, משליך ישירות על הסיבוכיות החישובית של כל איטרציה בתהליך הפענוח ; הגורם השני הינו מספר האיטרציות הנדרש על מנת להשיג את רמת הביצועים הרצויה. בעבודה זו, אנו מפתחים חסמים תחתונים אינפורמציוניים על כל אחד מגורמים אלו.

חלקו הראשון של המחקר, המוצג בפרקים 2 ו-3 של עבודה זו, מתמקד בגזירת חסמים עליונים על הקצבים ברי ההשגה של קודי בלוק ליניאריים ובינאריים וכן בפיתוח חסמים תחתונים על הסיבוכיות הגרפית של קודים אלו. חסמים אלו מתבססים על טענות מתורת האינפורמציה ותקפים עבור המקרה בו אורך הבלוק של הקודים שואף לאינסוף והפענוח מבוצע על ידי מפענח סבירות מרבית (ML). בפרק 2 של העבודה מוצגים חסמים עבור קודים המשודרים על גבי ערוצים חסרי זיכרון בינאריים במבוא וסימטריים ביציאה (MBIOS). גזירת חסמים אלו מבוססת של שיפור טכניקות חסימה שהוצגו על ידי Burshtein ושותפיו (לגבי קצבים ברי השגה) ועל ידי Sason ו-Urbanke (לגבי סיבוכיות גרפית). בעבודות קודמות אלו, האנליזה התבססה על קוונטיזציה בינארית של המידע המתקבל מערוץ התקשורת. שיפור החסמים בעבודה זו נעשה על ידי שיפור טכניקת החסימה כך שתאפשר כל קוונטיזציה סימטרית של המידע המתקבל מהערוץ ואף אנליזה ישירה של מידע זה ללא קוונטיזציה כלל. החסמים מופנים להערכה של הברירה בין הביצועים והסיבוכיות לאיטרציה של קודי LDPC תחת פענוח איטרטיבי. בפרק 3 של העבודה מוצגת הכללה של חסמים אלו למקרה בו הקודים משודרים על גבי ערוצי MBIOS מקביליים, כאשר כל אחת מסיביות הקוד מופנית לערוץ ספציפי. הכללה זו משמשת להערכת הברירה בין הביצועים והסיבוכיות לאיטרציית פענוח עבור קודים מנוקבים (punctured codes). השיפור בהדיקות החסמים לעומת חסמים ידועים מהספרות נבחן אף הוא. החסמים משמשים על מנת לאמוד את חלקו של הפער לקיבול הנגרם על ידי תת-האופטימליות של אלגוריתמי הפענוח האיטרטיביים היעילים ואת החלק הנובע מהמבנה האלגברי של הקודים.

בחלקו השני של המחקר, המוצג בפרק 4 של עבודה זו, אנו גוזרים חסמים תחתונים על מספר איטרציות הפענוח הנדרש על מנת להשיג רמת ביצועים רצויה, כאשר השידור הינו על גבי ערוץ מחיקה בינארי ואורך הבלוק של הקודים שואף לאינסוף. גזירת החסמים מתבססת על ניתוח פונקציות העברת אינפורמציה אקסטרינזית (EXIT functions) של מרכיבי הקוד השונים ומתבססת על כך שפונקציות אלו מהוות ייצוג מדויק של תהליך הפענוח האיטרטיבי תחת ההנחות לעיל. החסמים נתונים כפונקציה של הפער בין קצב הקוד לקיבול הערוץ, הינם פשוטים לחישוב ודורשים אך ורק ידיעה של אחוז צמתי המשתנים בגרף שדרגתם שווה לשתיים. אנו בוחנים צבירים של קודי LDPC וכן את המשפחות החדשות יותר של קודי חזרה-אגירה ווריאציות שלהן (כגון קודי accumulate-repeat-accumulate). לכל המשפחות המוזכרות לעיל, אנו מראים כי מספר האיטרציות הנדרש עבור פענוח מוצלח גדל לפחות ביחס הפוך לפער לקיבול; תוצאות אלו תואמות להשערה של Khandekar ו-McEliece משנת 2001 וכן לתוצאות ניסיוניות.

בחלקו האחרון של המחקר, המוצג בפרק 5 של עבודה זו, אנו פונים לבחינת מגבלות הביצועים של קודים בעלי אורך בלוק סופי כאשר התקשורת מתבצעת על גבי ערוץ סימטרי וחסר זיכרון. אנו מתמקדים בבחינת חסמים מסוג אריזת כדורים (sphere-packing bounds), שהינם חסמים תחתונים על הסתברות השגיאה המתבססים על ניתוח גיאומטרי של אזורי ההחלטה במפענח שרירותי של הקודים. חסמים אלו תקפים תחת פענוח ML וכן עבור פענוח רשימה (list decoding). חסמים תחתונים אלו משמשים באופן

תדיר כמדד על המרחק בין ביצועים של קוד נתון תחת אלגוריתם פענוח מעשי כלשהו ובין המגבלות התיאורטיות על הביצועים של קוד אופטימלי בעל אותו קצב ואורך בלוק המשודר על גבי אותו ערוץ. בגבול האסימפטוטי כאשר אורך הבלוק של הקודים שואף לאינסוף, החסם הידוע ההדוק ביותר על הסתברות השגיאה הינו חסם אריזת הכדורים שהוצא בשנת 1967 על ידי Shannon, Gallager ו-Berlekamp. חסם זה תקף עבור ערוצים דיסקרטיים וחסרי זיכרון ומאופיין בין השאר ע"י כך שההתנהגות האקספוננציאלית שלו מדויקת בתחום הקצבים שבין הקצב הקריטי של הערוץ והקיבול. האנליזה משנת 1967 התמקדה בהתנהגות המעריכית האסימפטוטית של החסם. הגישה שהוליכה את המחברים היתה לפשט ככל הניתן את הגזירה, כל עוד לא נפגעת ההתנהגות האסימפטוטית של החסם. גישה זו הביאה לחסם שאינו יעיל עבור קודים באורכי בלוק מעשיים. הביצועים המצוינים של קודים מודרניים, אשר משיגים קצבים הקרובים לקיבול הערוץ גם עם אורכי בלוק לא ארוכים מדי, הובילו בשנת 2004 את Valembois ו- Fossorier לבחון מחדש את טכניקת הגזירה של חסם אריזת הכדורים משנת 1967 ולהתאימה בצורה טובה יותר עבור קודים באורך בלוק סופי. הם הבחינו במספר נקודות במהלך הגזירה בהן ניתן היה לעדן את טכניקת החסימה ובכך לשפר את הדיקות החסם עבור קודים בעלי אורך בלוק סופי. כמו כן, שיפורים אלו אפשרו את הפעלת החסם עבור ערוצים בעלי אלפבית מוצא אינסופי ואף רציף. במחקר זה, אנו מציגים חסם אריזת כדורים משופר (ISP) אשר מהדק משמעותית את החסם של Valembois ו-Fossorier. חסם ה-ISP תקף עבור קודים המשודרים על גבי ערוצים סימטריים וחסרי זיכרון. גזירת חסם אריזת הכדורים מ-1967 וכן הגרסה המהודקת שלו נשענות על צעד ביניים, בו נחסמת הסתברות השגיאה המקסימלית של קודים בעלי הרכב קבוע (fixed composition). חסם זה משמש כבסיס לגזירת החסם על הסתברות השגיאה הממוצעת של קודים כללים, תוך הפסד משמעותי בהדיקות. בעבודה זו, אנו מראים שכאשר קיימת סימטריה בערוץ, ניתן לנצל תכונה זו על מנת לחסום ישירות את הסתברות השגיאה הממוצעת של קודים כללים. השיפור הניכר בהדיקות, בעיקר עבור קודים באורך קצר עד בינוני, מודגם על ידי תוצאות נומריות. כמו כן מודגם כי עבור אותות המאופננים בשיטת phase shift keying (PSK) ומשודרים על גבי הערוץ הגאוסי האדיטיבי (AWGN), חסם ה-ISP מהווה אלטרנטיבה מעניינת לחסם אריזת הכדורים הקלאסי של Shannon משנת 1959 המותאם לערוץ הגאוסי בלבד ואינו מניח הנחות אפריוריות לגבי סוג האפנון (מעבר להנחה כי האותות הינם שווי אנרגיה).