PROBLEMS IN CODED COMMUNICATIONS: PERFORMANCE BOUNDS AND POLAR CODING

ERAN HOF

PROBLEMS IN CODED COMMUNICATIONS: PERFORMANCE BOUNDS AND POLAR CODING

RESEARCH THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

ERAN HOF

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

Elul 5770 HAIFA SEPTEMBER 2010

THIS RESEARCH THESIS WAS SUPERVISED BY PROF. IGAL SASON AND PROF. SHLOMO SHAMAI UNDER THE AUSPICES OF THE DEPARTMENT OF ELECTRICAL ENGINEERING

ACKNOWLEDGMENT

I wish to thank my advisors Prof. Igal Sason and Prof. Shlomo Shamai for their dedicated supervision, to my parents, Lea and Abraham who raised and educated me with love and care, to numerous teachers, friends, and colleges from which I learned valuable lessons and skills, and to my newlywed wife Irit for her support and companionship.

The generous financial help of the Technion is gratefully acknowledged.
This research was supported by the Israel Science Foundation (grant no. 1070/07) and by the European Commission in the framework of the FP7 Network of Excellence in Wireless Communications NEWCOM++.

Contents

List of Figures

List of Tables

Abstract

Our study begins with the error performance analysis of non-binary codes. The performance of non-binary linear block codes is studied via the derivation of new upper bounds on the block error probability under maximum-likelihood decoding. The transmission of these codes is assumed to take place over a memoryless and symmetric channel. The new bounds, which are based on the Gallager bounding technique and their variations, are applied to regular ensembles of non-binary low-density parity-check codes. These upper bounds are also compared with sphere-packing lower bounds. Our study indicates that the new upper bounds are useful for the performance evaluation of coded communication systems which incorporate non-binary coding techniques.

Secondly, erasure and list decoding of linear block codes are concerned. A message independence property and some new upper bounds on the performance are derived for erasure, list and decision-feedback schemes with linear block codes transmitted over memoryless symmetric channels. Similar to the classical work of Forney, we focused on the derivation of some Gallager-type bounds on the achievable tradeoffs for these coding schemes, where the main novelty is the suitability of the bounds for both random and structured linear block codes (or ensembles). The bounds are applicable to finite-length codes and the asymptotic case of infinite block length, and they are applied to low-density parity-check code ensembles.

Next, a modified Viterbi algorithm with erasures and list-decoding is introduced. This algorithm is shown to yield the optimal decoding rule of Forney with erasures and variable list-size. For the case of decoding with erasures, the optimal algorithm is compared to the simple algorithm of Yamamoto and Itoh. The comparison shows a remarkable similarity in simulated performance, but with a considerably reduced decoding complexity.

Finally, two applications for the method of channel polarization are studied. Polar coding, recently introduced by Arikan, is a structured coding technique which is shown to approach capacity for every output-symmetric discrete memoryless channel. The theory of polar coding is still in its early days. Consequently, no known polar coding

schemes have been shown to compete well with the state of the art of other modern coding schemes. Nevertheless, it has already been shown that channel polarization techniques may be applied for various multi-user information-theoretic problems. The application of channel polarization is investigated in our research for two different communication problems: signaling over parallel channels, and secure communication over the wire-tap channel.

Notation

- x – Scalar.
- $\mathbf{x}$ – Row vector.
- $\mathcal{X}$ – Set.
- $\mathsf{E}(X)$ – Expectation of X.
- $\Pr(E)$ – Probability of E.
- $\emptyset$ – Empty set.

Abbreviations

- ARQ – Automatic repeat request
- AWGN – Additive white Gaussian noise
- BCJR – Bahl, Cocke, Jelinek, and Raviv
- BICM - Binary interleaved coded modulation
- BPSK – Binary phase shift keying
- BSC – Binary symmetric channel
- CSI – Casual state information
- DMC – discrete memoryless channel
- DS2 – Duman and Salehi (second version)
- GRS - Generalized Reed-Solomon
- GSM – Global System for Mobile communications
- ISP – Improved sphere-packing
- LDPC – Low-density parity-check
- LR – Likelihood-ratio
- MBIOS – Memoryless binary-input output-symmetric
- ML – Maximum likelihood
- RS – Reed-Solomon
- SFB - The Shulman and Feder bound
- SISO – Soft-in Soft-out

- SP59 – The 1959 sphere-packing bound
- SPC – Single parity-check.
- VA – Viterbi Algorithm
- YI – Yamamoto-Itho

Chapter 1

Introduction

The theoretical foundations of communication theory were laid at the mid of the twentieth century by the celebrated paper of C. E. Shannon [98]. Two dual problems are conceived and solved in this paper: the problem of *source coding*, and the problem of *channel coding*. The goal of the channel coding problem is to achieve reliable communication at the maximal rate over a noisy channel. This goal is satisfied by introducing some redundancy to the transmitted sequence. This operation is carried by the *channel encoder*. The redundancy in the transmitted signal supports the decoding of the transmitted sequence. This dissertation is focused solely on the channel coding problem.

One of the most fascinating results introduced by Shannon is that information can be communicated with arbitrarily small distortion at rates arbitrarily close to capacity. This result completely contradicts all that could have been intuitively understood from the state of the art of the communication theory and practice of Shannon's time. Shannon's solution to the channel coding problem relies on using *random block codes*. This technique is referred to in the literature as the *random coding* technique, and it serves as one of the fundamental tools of information theory. Nevertheless, fully-random block codes are of little practical interest (mainly due to complexity and delay concerns). The pursuit after practical coding schemes which reliably operate close to the channel capacity limit, contributed to more than 60 years of research in coding theory. Information and coding theorists spread their interest between the following core subjects:

- Understanding the fundamental limits in coded communications.
- Formulation and adaptation of coding schemes to support various communication models and achieving their fundamental limits.

- Development and analysis of algorithmic techniques operating over the formulated coding schemes.

1.1 A Glimpse to Channel Coding

A short introduction to the theory of channel coding is provided. The modest goal of this introduction is merely to mention the coding schemes, terms, and techniques that are of some interest in the following chapters.

Consider the problem of reliable transmission of digital information over a noisy channel. The channel encoder introduces some redundancy to the digital information before its transmission over a noisy channel. This redundancy supports the decoding process at the receiver. The set of possible coded sequences, generated by the channel encoder, is called the *channel code* or the the *codebook*. The term *error-correcting code* (or *forward error-correcting code*) is very common, as one of the key purposes of channel codes is to provide the means of error correction capabilities. Some codes support the less demanding purpose of *error detection*, in order to facilitate automatic mechanisms for repeat requests.

Much of the interest in coding theory is devoted to *linear* codes. These codes may facilitate substantial algebraic structures, while still maintaining the potential of achieving reliable communications at rates arbitrarily close to channel capacity (see, e.g., [46, Section 6.2] and [111, Section 3.10]). Linear block codes can be represented by a *generator matrix* whose rows form the basis vectors of the linear code. Alternatively, the code may be defined by a *parity-check matrix* whose rows form a basis of the vector space which is orthogonal to the code. The algebraic structure of linear codes allows the introduction of practical encoding and decoding algorithms.

Early influential examples of linear block codes include the well-known codes of Golay [47] and Hamming [49]. Other important families of linear block codes are the Bose-Chaudhuri-Hocquenghem (BCH), Reed-Solomon (RS), and generalized RS (GRS) codes (see, e.g., [15], [71], [89] and references therein). These codes possess elegant and sophisticated algebraic structures. Hence, these codes are referred to in the coding literature as *algebraic codes*. The field of algebraic coding theory contributed to many of the successes of coding theory in its early decades. Algebraic codes are an immanent part of many important applications in both communication systems and storage.

Convolutional codes, invented by Elias in 1955, are one of the key machineries of communication systems [35]. These codes have a linear structure, and can be

described by a discrete time finite-state machine. Based on their tree structure, convolutional codes can be decoded using sequential decoding algorithms [38], [118]. The recognition of the practical usage of convolutional codes was further increased with the introduction of Massey's threshold decoding algorithms (and its variations) [76]. Convolutional codes possess the pleasing feature of having a *trellis-graph* structure (see, e.g., [101]). The trellis structure enables the introduction of practical and optimal decoding algorithms. For detailed description on the structure and techniques related to convolutional coding, see [60], [71] and references therein. The vast spread of convolutional codes in practice is due to two important contributions: The Viterbi algorithm (VA) and concatenated coding schemes. The VA was introduced by Viterbi in 1967 [42], [108]. The algorithm is popular in many coding and signal processing applications, as it yields the maximum-likelihood (ML) solution while being amenable to practical implementations. Concatenated coding, introduced by Forney in 1966 [43], is a coding technique incorporating two relatively short codes which are combined to construct an efficient and strong (relatively long) code. The serially concatenated scheme of a convolutional codes with an RS code is one of the most popular coding schemes in the pre Turbo-era (see [22], [71], and references therein).

Turbo coding, introduced by Berrou, Glavieux, and Thitimajshima in 1993, is the first channel coding scheme demonstrated to operate reliably over the Additive White Gaussian Noise (AWGN) channel within 1 dB from capacity [13]. The original turbo structure comprises a *parallel concatenation* of convolutional component codes where one of the code is an interleaved version of the other. This structure allows to construct structured but *random-like* codes. The turbo decoding algorithm is based on *iterative* soft-decoding of each of the code components. Each component is decoded based on the BCJR algorithm, introduced in 1974 by Bahl, Cocke, Jelinek, and Raviv [7]. The BCJR algorithm allows for optimal symbol-wise decoding of codes which possess a trellis structure. Being a *soft-in soft-out* (SISO) algorithm, the BCJR algorithm provides soft reliability information on each of the decoded symbols based on soft a-priori reliability inputs on these symbols. The turbo principle is based on iterative exchanging and refinement of these soft values between the two SISO decoding algorithms of each of the two component codes.

The discovery of turbo codes started the modern era of the channel coding theory. Among its important contributions is the rediscovery of Gallager's Low-density parity-check (LDPC) codes [44]. LDPC codes are linear block codes, which possess a *sparse* structure. The sparse structure of LDPC codes supports the use of practical iterative decoding algorithms. Various structured LDPC and other *turbo-like* codes and their related iterative decoding techniques were reported in the last decade. Many of these

channel coding techniques show remarkable performance near the ultimate channel capacity limit with tolerable complexity. For a detailed study of modern coding theory, the reader is referred to [86], [90], and references therein.

Channel codes for bandwidth-limited channels comprise a valuable part of coding theory and practice. Important families of non-binary codes for communicating over bandwidth-limited channels are based on lattices [21], trellis-coded modulation [14], [106], multilevel coding [59], and bit-interleaved coded modulation [2], [121]. Modern coding schemes and techniques were also incorporated to construct spectral-efficient schemes (see, e.g., [10], [11], [12], [25], [33], [72], [85], [87], [112], [113], and references therein).

1.2 Performance Analysis of Coded Communication Systems

A substantial part of coding theory is dedicated to the performance analysis of coded communication systems. The analysis of error performance is of particular interest where both the fundamental limitations on the decoding error probability in general, and the error performance of structured coding schemes are studied.

Error performance characteristics of coded communication systems rarely admit exact closed-form expressions. Consequently, the performance of these systems is usually analyzed via upper and lower bounds on the decoding error probability. Modern coding schemes perform reliably at rates which are close to the channel capacity, whereas union bounds are useless for codes of moderate to large block lengths at rates above the channel cut-off rate. The limitation of the union bound therefore motivates the introduction of some improved bounding techniques which can also be efficiently calculated. Although the performance analysis of specific codes is in general prohibitively complex, this kind of analysis is tractable for various code ensembles for which the derivation of some of their basic features (e.g., the average distance spectrum) lends itself to analysis.

1.2.1 Error Performance under Maximum-Likelihood (ML) Decoding

The 1965 Gallager bound [45] is one of the well-known upper bounds on the decoding error probability of ensembles of fully random block codes, and it is informative at all rates below the channel capacity limit. Emerging from this bounding technique, the bounds of Duman and Salehi (see [31] and [32]) possess the pleasing feature that

they are amenable to analysis for codes or ensembles for which the (average) distance spectra are available.

The bounds of Duman and Salehi, in particular its second version (called hereafter the 'DS2 bound'), are generalized in [26], [94], and [96] for various memoryless communication systems. Moreover, the DS2 bound facilitates the derivation of a large class of previously reported bounds (or their Chernoff versions), as shown in [94] and [96]. Gallager-based bounds for binary linear block codes whose communication takes place over fading channels are provided in [58], [93] and [119]. The Shulman and Feder bound (SFB) [100] forms an extension of the 1965 Gallager bound which can be also applied to structured codes or ensembles. An adaptation of the SFB to non-binary linear block codes was reported in [11] for the case of coding with a random coset mechanism (see, e.g., [11], [12], [46], and [57]), and for the case of transmission over modulo-additive noise channels (see [37]). Generalizations of Gallager-type bounds, among them the DS2 bound, for the case of binary linear block codes whose transmission take places over parallel channels are provided in [73] and [92].

The 1959 sphere-packing (SP59) bound of Shannon [99] is a lower bound on the decoding error probability of block codes whose transmission takes place over the additive white Gaussian noise (AWGN) channel with equal-energy signaling. The 1967 sphere-packing bound of Shannon, Gallager and Berlekamp [97] forms an alternative lower bound on the decoding error probability of block codes which applies to discrete memoryless channels. An improved sphere-packing (ISP) bound, which holds for all memoryless symmetric channels, was recently derived in [115] by improving the bounds in [97] and [107].

For a comprehensive tutorial on the performance analysis of binary linear block codes under maximum-likelihood (ML) decoding, the reader is referred to [94] and references therein.

1.2.2 Error Performance under Generalized Decoding

Exponential error bounds for the fully-random block code ensemble were derived and studied by Forney [41], referring to the following two situations:

1. A decoder is allowed not to make a decision on a received signal, or rejecting all estimates; this output is called an *erasure*. When a decision is made, the event where the decoder decision is incorrect is called an *undetected error*.

2. A decoder is allowed to make more than one estimate of the received signal. The output of this decoder forms a list of codewords, and the event where the transmitted message is not on the list is called a *list error event*.

Following [109], decoding rules for these two situations are called in the following parts of this dissertation *generalized decoding rules*. As explained in [41], erasure and list options may be useful when the transmitted data contains some redundancy, when a feedback channel is available, or when several stages of coding (e.g., concatenated codes) are used. The size of the decoded list in [41] is allowed to vary according to the received signal. This decoding rule differs from [36] and [117] where the size of the list is predetermined and fixed.

Consider the case of decoding with erasures (the first situation). By allowing a decoder to increase the probability of erasures, the undetected error probability can be reduced. In the case of list decoding (the second situation), by increasing the decoder list, the list error probability can be reduced. The optimum decoding rules with respect to these tradeoffs were provided in [41] and they were analyzed via the derivation of exponential bounds for the fully-random block code ensemble. Sub-optimal decoding rules are analyzed in [9], [52], and [54], via a similar bounding technique, and the random coding error exponents under optimal and sub-optimal decoding rules are compared. It is noted that the considered decoding rules are studied with respect to a given code, and finding the optimal codes for these scenarios remains an open problem.

The performance analysis under generalized decoding rules with erasures enables the study of coded communications with a noiseless decision feedback. Specifically, it is assumed that erasures are followed by a repeat-request acknowledgment over a noiseless and immediate feedback channel. Such schemes are often referred to as hybrid automatic repeat request (ARQ) systems. Unlike the channel capacity of single-user discrete memoryless channels (DMC), which is not affected by feedback (see for example [23, p. 216]), a significant improvement is demonstrated in [41] for the error exponents of the concerned coded schemes. In this respect, the reader is also referred to [48] where the error exponents of hybrid ARQ schemes with limited retransmissions are studied. The effect of feedback was also considered in [19], and it was shown to significantly reduce the block error probability for DMCs.

1.3 Polar Coding

Channel coding via the method of *channel polarization* was recently provided by Arikan in [4]. On a binary-input DMC, polarization ends up with either 'good bits', i.e., binary channels whose capacity approaches 1 bit per channel use, or 'wasted bits', i.e., channels whose capacity approaches zero. The fraction of the good bits is equal to the mutual information with equiprobable inputs (which equals the capacity for

the case of a memoryless symmetric channel). The rate of channel polarization and additional characterization of polar codes were studied in [6], [63] [64] and [91].

For a single-user channel-coding problem, the polar coding scheme is based on transmitting the uncoded information bits over the capacity approaching channels (when we interpret the polarization as a kind of a precoding or pre-processing). At the same time, fixed and predetermined bits are transmitted over the channels whose capacity approaches zero. These predetermined bits are essential part of the successive decoding process of polar codes. Hence, these fixed and predetermined bits should not be ignored. In a physically degraded setting, as mentioned in [4], an order of polarization is maintained in the sense that 'good' bits for the degraded channel, must also be 'good' for the better channel.

1.4 Motivation and Related Work

1.4.1 Error Performance of Non-Binary Codes under ML Decoding

The drawback of the union bound (for codes of moderate to large block lengths the union bound diverges above the channel cut-off rate) motivates the derivation of upper bounds on the decoding error probability of non-binary codes. In particular, the derivation of Gallager-type bounds for non-binary linear block codes (or code ensembles) whose transmission takes place over memoryless symmetric channels are considered.

The definition of symmetry for channels whose input is non-binary should generalize the common definition of memoryless binary-input output-symmetric (MBIOS) channels. It is well known that for MBIOS channels, the decoding error probability under ML decoding is independent of the transmitted message (see, e.g., [111]). Many of the bounding techniques for binary linear block codes under ML decoding are based on this message independence property. Hence, the motivation for investigation of the possible generalization of this result for non-binary codes is clear.

The study of non-binary LDPC code ensembles further motivates the derivation of the suggested bounds. The performance analysis of binary LDPC ensembles in [44] is carried under the assumption that the channel is MBIOS. In contrast to the binary case, the performance analysis provided in [44] for non-binary LDPC code ensembles is carried under a symmetry assumption which is tailored to the specific bounding technique that was introduced in [44]. The asymptotic error performance of several non-binary LDPC structures is studied in [11] under ML decoding. Their asymptotic

performance under iterative decoding was studied in [12], and further bounds on the thresholds of non-binary LDPC code ensembles were studied in [85] and [113]. It was assumed in [11] that the transmission takes place over channels with a random coset mechanism which enables to dismiss the channel symmetry condition required in [44]. The decoding error probability of various non-binary LDPC code constructions was studied empirically in the literature (see, e.g., [25]).

1.4.2 Error Performance of Structured Codes under Generalized Decoding

Much of the current literature on performance analysis of coded communications is focused on maximum-likelihood (ML) decoding (see, e.g., [94] and references therein). Lower bounds on the error exponents for fully-random block codes under generalized decoding rules were derived in [9], [41], [77], and [109]. Error exponents are provided in [103] and [104] for random codes with a constant composition under some suboptimal decoding rules. An upper bound on the error exponent under fixed-size list-decoding was provided in [97]. The error performance under fixed-size list-decoding was studied for specific codes in [8], [17] and [68] where the communication takes place over an AWGN channel. Additional (suboptimal) decoding rules with erasures were analyzed in [28] and [29].

Consider the case where a given structured code (or code ensemble) is transmitted over a DMC. A vast amount of performance analysis techniques are available under ML decoding. On the other hand, for the case of generalized decoding (either optimal a'la Forney or other suboptimal decoding rules) only few general analysis techniques exist. This gap motivates the study of a possible adaptation of analysis techniques for the case of generalized decoding. Specifically, upper bounds on the error probability under generalized decoding algorithms are of interest.

As mentioned in Section 1.4.1, many of the bounding techniques under ML decoding rely on a message independence property. Specifically, the error probability of binary linear block codes whose transmission takes place over MBIOS channels is known to be independent of the transmitted codeword. Hence, the motivation for the study of corresponding message independence properties under generalized decoding rules emerges.

1.4.3 Optimal Erasure and List Decoding of Convolutional Codes

The VA is known to yield the ML sequence estimation for a finite-state Markov process that is observed via a memoryless channel. For the particular case of coded communications with convolutional codes, the output of the VA coincides with the ML decision. The simplicity and low complexity of the VA lead to its wide spread as one of the key techniques in communication theory and practice (see, e.g., [110], [22], and references therein).

Many generalizations and adaptations of the VA were reported in the literature, both in coding theory and signal processing. In this dissertation, applications of list decoding and ARQ schemes are concerned. List decoding generalizations of the VA were reported in [20], [50], [70], [81], [88], [95], and [102]. Adaptations of the VA to support hybrid ARQ schemes were reported in [51], [53], [67], and [120]. These results motivate the pursuit for a variation on the VA, where optimal decoding a'la Forney is considered. Specifically, a feasible implementation of Forney's optimal decoding rule with erasures (and possible variable list-size) is studied for the case of convolutional codes.

1.4.4 Applications of Polar Codes

The theory and practice in polar coding is still in its early days. Nevertheless, the method of channel polarization and its related techniques have already been introduced in several important problems of information theory. Applications of polar codes for basic multi-terminal models such as the degraded broadcast channel and the multiple-access channel, were studied in [62]. Polar codes were also found to be optimal for lossy source coding [62], [65]. Applications of polar coding to binary Wyner-Ziv and the binary Gelfand-Pinsker problems were provided in [65]. The compound capacity of polarization codes (under successive cancelation decoding) was studied in [55]. The variety of these possible applications motivates the study of further applications where the channel polarization technique can be used.

1.5 This Dissertation

Chapter 2 is focused on the performance analysis of non-binary linear block codes under ML decoding. A definition of symmetry is stated for memoryless channels with non-binary input alphabets. Under the considered symmetry condition, it is proved that the conditional error probability under ML decoding is independent of

the transmitted codeword. This result generalizes the well-known message independence property for MBIOS channels. Moreover, this result is in agreement with [39] and [40] which prove the same result under linear-programming decoding. The rest of Chapter 2 is devoted to the derivation of upper bounds on the error performance of non-binary linear block codes under ML decoding. The upper bounds on the error performance derived in this chapter are applied to non-binary regular LDPC code ensembles of Gallager [44], and their error performance is studied for various communication channel models. The exact complete composition spectra for these LDPC code ensembles are also provided (instead of the upper bound in [44]), and this exact analysis forms a generalization of the analysis in [18] and [105]. In addition, the derived upper bounds are compared to sphere-packing lower bounds on the decoding error probability for various code ensembles.

Chapter 3 considers upper bounds on the error probabilities under generalized decoding rules. The provided bounds are suitable for linear block codes whose transmission takes place over memoryless symmetric channels. These bounds are accompanied by some message-independence results for the considered generalized decoding rules. Both optimal and suboptimal decoding rules are considered. When variable-size list-decoding is considered, upper bounds on the expected size of the decoded list and the associated error probability under list decoding are jointly derived. In addition, upper bounds on the list error probability are introduced for linear block codes where the size of the list is fixed. The bounds derived in this chapter are applicable to the performance analysis of specific codes and ensembles via their (average) distance spectra. The bounds are suitable for finite block lengths and also for asymptotic analysis. The provided results are exemplified for two coding schemes: Fully-random linear block codes, and regular binary and non-binary LDPC code ensembles with finite block lengths. Applications of the provided bounds for the study of hybrid-ARQ schemes are also exemplified.

Chapter 4 studies some generalizations of the VA for list-decoding and decoding with erasures. A modification of the VA is introduced, which coincides with the optimal decoding rule of Forney for the cases at hand. Although presented for the decoding of convolutional codes, the provided algorithm is suitable for the more general case where finite-state Markov processes are observed via memoryless channels. The simulated performance of the proposed modification is compared in this chapter with the simulated performance of two suboptimal decoding algorithms with erasures: the likelihood-ratio (LR) test decoding rule, and a simple decoding scheme with repeat requests that was introduced by Yamamoto and Itoh in [120]. Even though the

decoding scheme in [120] is remarkably simple, the comparison shows good similarity between the performance of the simple scheme to the optimal one. On the other hand, the performance of the decoding algorithm based on the LR test is considerably degraded in comparison with that of the optimal performance.

In Chapter 5 channel polarization is applied for the wire-tap communication model. The wire-tap channel model is a multi-user communication model involving a single transmitter, and two receivers: one of the legitimate-user and one of the eavesdropper. The maximal rate under which secure and reliable communication is possible is called the secrecy capacity. Our study shows that the secrecy capacity can be achieved by proper application of the channel polarization method for degraded and symmetric channels. It is proved that for every rate below the channel secrecy capacity, there exists a suitable polar code for which reliable and secure communication are achieved under successive cancelation decoding. Moreover, our polar coding scheme is shown to achieve the entire rate-equivocation region for the considered channel model.

In Chapter 6 channel polarization is applied for signaling over parallel channels. It is shown that using the method of channel polarization, the capacity of signaling over arbitrarily-permuted memoryless and symmetric parallel-channels is achievable under an assumption of channel degradation. A channel coding scheme and its corresponding successive cancelation decoding algorithm are proposed. The proposed scheme incorporates the method of channel polarization with an algebraic maximum-distance separable codes. The achievable rates of the provided scheme are also studied in the general case where the assumption on channel degradation is removed.

Chapter 2

Performance Bounds for Non-Binary Linear Block Codes over Memoryless Symmetric Channels

Chapter Overview

The performance of non-binary linear block codes is studied in this chapter via the derivation of new upper bounds on the block error probability under ML decoding. The transmission of these codes is assumed to take place over a memoryless and symmetric channel. The new bounds, which are based on the Gallager bounds and their variations, are applied to the Gallager ensembles of non-binary and regular LDPC codes. These upper bounds are also compared with sphere-packing lower bounds. This study indicates that the new upper bounds are useful for the performance evaluation of coded communication systems which incorporate non-binary coding techniques.

The general concept used in this chapter is based on a partitioning of the original ensemble into two subsets of codebooks according to their minimal Hamming distance. For the set of codebooks whose minimal distances are below a certain value (which is later determined in order to achieve a tight bound), a simple union bound is used which only depends on their distance properties. As for the complementary set of codebooks (whose minimal Hamming distance is larger than the above value), a Gallager-type bound on the decoding error probability is used; the latter bound depends both on the distance properties of the ensemble and the communication

channel, and it relies on a generalization of the DS2 bound to non-binary linear block code ensembles. The chapter is based on the following paper:
E. Hof, I. Sason, and S. Shamai (Shitz), "Performance bounds for non-binary linear block codes over memoryless symmetric channels," *IEEE Trans. on Information Theory*, vol. 55, no. 3, pp. 977–996, March 2009.

This chapter is structured as follows: the symmetry requirements and the message independence proposition are provided in Section 2.1. The proposed bounding approach is introduced in Section 2.2, and these bounds are exemplified for the Gallager LDPC code ensembles over a q-ary symmetric and AWGN channels. Variations of these bounds are also derived and exemplified in Section 2.3 for fully-interleaved fading channels with perfect casual state information (CSI) at the receiver. Section 2.4 concludes the discussion. Various technical details are relegated to the appendices.

2.1 Channel Symmetry and Message Independence

Let $\mathcal{X} = \{x_0, x_1, \ldots, x_{q-1}\}$ be a given alphabet with cardinality q. We assume an addition operation $(+)$ over the alphabet $\mathcal{X}$ for which $\{\mathcal{X}, +\}$ forms an Abelian group. Let $x_0 = 0$ be the additive identity of this group. In addition, let $\mathcal{Y}$ be a given discrete (or continuous) alphabet. We assume a memoryless channel, and denote the channel transition probability (or probability density, respectively) function by $p(y|x)$, where $x \in \mathcal{X}$ and $y \in \mathcal{Y}$.

Definition 2.1 (Channel symmetry) A memoryless channel which is characterized by a transition probability p, an input-alphabet $\mathcal{X}$ and a discrete output alphabet $\mathcal{Y}$ is *symmetric* if there exists a function $\mathcal{T} : \mathcal{Y} \times \mathcal{X} \to \mathcal{Y}$ which satisfies the following properties:

1. For every $x \in \mathcal{X}$, the function $\mathcal{T}(\cdot, x) : \mathcal{Y} \to \mathcal{Y}$ is bijective.
2. For every $x_1, x_2 \in \mathcal{X}$ and $y \in \mathcal{Y}$, the following equality holds:

$$p(y|x_1) = p(\mathcal{T}(y, x_2 - x_1)|x_2). \tag{2.1}$$

Remark 2.1 For channels whose output alphabet is continuous, an additional requirement on the mapping $\mathcal{T}$ is that its Jacobian is equal to 1.[1] In this case, the

[1] It is possible to use a generalized definition for both discrete and continuous output alphabets using the notion of unitary functions as done for example in [115, Section III-A].

condition in (2.1) implies that

$$\int p(y|x_1)\, dy = \int p(\mathcal{T}(y, x_2 - x_1)|x_2)\, dy.$$

Example 2.1 (MBIOS channels) For the particular case of channels with a binary-input alphabet, and whose output alphabet $\mathcal{Y}$ is the set of real numbers, setting

$$\mathcal{T}(y, x) = \begin{cases} y & \text{if } x = 0 \\ -y & \text{if } x = 1 \end{cases}$$

then Definition 2.1 coincides with the standard definition of MBIOS channels. The meaning of the function $\mathcal{T}$ is better understood via the setting of MBIOS channels. Referring to (2.1), the transition probability given a channel input x_1 is equal to the transition probability given another input x_2 where the sign of the output is changed if the two binary inputs are different.

Example 2.2 (Random coset mechanism followed by an arbitrary channel) In [11], [46] and [57], the transmission of block codes takes place over an arbitrary memoryless channel followed by a random coset mechanism. That is, instead of transmitting the coded message $\mathbf{x}$, the vector $\mathbf{x}+\mathbf{v}$ is transmitted where $\mathbf{v}$ is a random vector and the addition is carried out symbol-wise. The random vector $\mathbf{v}$ is called the coset, and it is known to both the transmitter and the receiver. When coding schemes with a random coset mechanism are applied to an arbitrary memoryless channel, the symmetry of the equivalent channel is guaranteed. To see this, consider the equivalent channel that includes the addition of the coset symbols followed by the original channel, and whose observations are pairs (y, v), where v is the random coset symbol added to the transmitted coded symbol, and y is the (original) channel output. Assuming a memoryless channel, the symmetry is guaranteed by setting

$$\mathcal{T}((y, v), x) = (y, v - x), \quad y \in \mathcal{Y},\ x, v \in \mathcal{X}$$

where $\mathcal{X}$ and $\mathcal{Y}$ are the input and output alphabets, respectively. Notice that $\mathcal{T}$ is now defined over $(\mathcal{Y} \times \mathcal{X}) \times \mathcal{X}$, where $\mathcal{Y} \times \mathcal{X}$ forms the output alphabet of the equivalent channel.

Based on Definition 2.1, we get the following lemma:

Lemma 2.1 let x_1, x_2, x_3 be arbitrary symbols in $\mathcal{X}$, and let p be a transition probability law of a memoryless symmetric channel. Then,

$$p\Big(\mathcal{T}\big(\mathcal{T}(y, x_1), x_2\big)|x_3\Big) = p\big(\mathcal{T}(y, x_1 + x_2)|x_3\big) \tag{2.2}$$

where $\mathcal{T}$ is the mapping satisfying the symmetry properties in Definition 2.1.

Proof: See Appendix 2.A. ■

For MBIOS channels, the capacity is attained with a uniform input distribution. In addition, random coding with a uniform (and memoryless) distribution attains the optimum random-coding error exponent provided by Gallager (see [45], [46], and [111]). The following lemma generalizes these results for the case of discrete, memoryless, and symmetric channels according to Definition 2.1 (a similar result follows for the case of memoryless symmetric channels with continuous output-alphabets).

Lemma 2.2 Let Q be a probability function over the input alphabet $\mathcal{X}$, and let p be a transition probability function of a discrete symmetric and memoryless channel. Then, the mutual information $I(Q)$, between the channel input (with an input probability distribution Q) and the channel output, given by

$$I(Q) = \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}} Q(x)p(y|x) \ln\left(\frac{p(y|x)}{\sum_{x'\in\mathcal{X}} Q(x')p(y|x')}\right)$$

and the Gallager function $E_0(\rho, Q)$ [46], defined by

$$E_0(\rho, Q) \triangleq -\ln\left(\sum_{y\in\mathcal{Y}}\left(\sum_{x\in\mathcal{X}} Q(x)p(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}\right),\ \rho \geq 0$$

are maximized (for every $\rho \geq 0$) by a uniform distribution.

Proof: The proof follows trivially by applying [111, Theorems 3.2.2 and 3.2.3] to the case at hand. ■

Lemma 2.2 is also valid for symmetric DMCs in the sense defined by Gallager in [46, p. 94] (as shown in the following definition):

Definition 2.2 (Gallager's definition for symmetric DMC [46]) A DMC is defined to be symmetric if the set of outputs can be partitioned into subsets in such a way that for each subset the matrix of transition probabilities (using inputs as rows and outputs of the subset as columns) has the property that each row is a permutation of each other row and each column (if more than 1) is a permutation of each other column.

Remark 2.2 It is easily verified that a symmetric DMC according to Definition 2.1, is symmetric according to Definition 2.2 of Gallager.

Consider linear block codes over the non-binary alphabet $\mathcal{X}$. Specifically, let $\mathbf{G}$ be a $k \times n$ matrix with components over the alphabet $\mathcal{X}$. Then, the linear block code with a generator matrix $\mathbf{G}$, denoted by $\mathcal{C} = \{\mathbf{x}_m\}_{m=1}^{q^k}$ where $\mathbf{x}_m = (x_{m,1}, \ldots, x_{m,n})$, is the set of all q^k linear combinations of the rows of $\mathbf{G}$. The conditional error probability of the m-th message is given according to

$$P_{\mathrm{e}|m} = \sum_{\mathbf{y} \in \Lambda_m^{\mathrm{c}}} p(\mathbf{y}|\mathbf{x}_m)$$

where Λ_m forms the decision region for the m-th codeword, and the superscript 'c' stands for the complementary set. The decision region of the m-th codeword under ML decoding gets the form

$$\Lambda_m = \left\{\mathbf{y} : p(\mathbf{y}|\mathbf{x}_m) > p(\mathbf{y}|\mathbf{x}_{m'}),\ \forall\, m' \neq m\right\}$$

and ties are resolved randomly with equal probability. A well-known result for binary linear block codes operating over MBIOS channels is that their error probability under ML decoding is independent of the actual transmitted codeword. This result enables a great simplification to the error performance analysis by assuming that the all-zero codeword, designated by $\mathbf{0}$, is transmitted. The following proposition is a generalization of this result for linear block codes communicated over memoryless and symmetric channels whose input alphabet is discrete (for the case of linear-programming decoding, see [40]):

Proposition 2.1 (Independence of the Conditional Error Probability on the Transmitted Codeword for all Memoryless Symmetric Channels) Let $\mathcal{C}$ be a linear block code whose transmission takes place over a memoryless and symmetric channel according to Definition 2.1. Then, the block error probability under ML decoding is independent of the transmitted codeword.

Proof: See Appendix 2.B. ■

The proof for the message independence property remains valid even if the channel transition probability is different for each transmission. This enables the analysis in Section 2.3 of q-ary PSK systems whose transmission takes place over fading channels with perfect CSI at the transmitter. In addition, note that in contrast to Lemma 2.2, Proposition 2.1 does not necessarily hold for symmetric DMCs in the broader sense, as in Definition 2.2 due to Gallager. This is demonstrated in the following counter-example:

Example 2.3 (Channel symmetry according to Definition 2.2 doesn't imply symmetry according to Definition 2.1) Consider a DMC with the integer ring $\mathbb{Z}_4$ (with arithmetic operations modulo-4) as common input and output alphabets, and with the following transition probability matrix:

$$P = [p_{i,j}] = \begin{pmatrix} 0.20 & 0.24 & 0.30 & 0.26 \\ 0.30 & 0.20 & 0.26 & 0.24 \\ 0.24 & 0.26 & 0.20 & 0.30 \\ 0.26 & 0.30 & 0.24 & 0.20 \end{pmatrix}.$$

In this matrix, the element $p_{i,j}$ (where $i, j \in \{1, \ldots, 4\}$) refers to the transition probability when the channel input is equal to $i-1$ and the output is equal to $j-1$. The memoryless channel which corresponds to P is symmetric according to Definition 2.2 (notice that each row and column is a permutation of another row or column, respectively). However, if the linear block code $\{00, 13, 22, 31\}$ is transmitted over the considered channel, then the resulting conditional error probabilities under ML decoding are 0.7540, 0.7210, 0.5424 and 0.7210, respectively, and they therefore depend on the transmitted codeword. To show this, we first need to determine the ML decoding regions for the considered code and channel. This is accomplished by evaluating the conditional probabilities of each possible output pair given each possible transmitted codeword (e.g., $p(03|31) = 0.26 \cdot 0.24 = 0.0624$). The decoding region for the all-zero codeword 00 is the set $\{22, 23, 32\}$ (note that the '00' vector is not included in the decision region of this codeword, and on the other hand, the vector '22' which forms a codeword is included in the decision region of the all-zero codeword). The conditional error probability given that the all-zero codeword is transmitted is therefore equal to $1 - p(22|00) - p(23|00) - p(32|00) = 1 - 0.30^2 - 0.30 \cdot 0.26 - 0.26 \cdot 0.30 = 0.7540$. The rest of the conditional error probabilities are similarly evaluated. Hence, due to Proposition 2.1, this channel is not symmetric according to Definition 2.1 although it is symmetric according to Definition 2.2.

2.2 Gallager Bounds for Memoryless Symmetric Channels and Some Applications

2.2.1 The DS2 bound

Let $\mathcal{C}$ be an (n, k) linear block code defined over the input-alphabet $\mathcal{X}$ with cardinality q. Consider the conditional error probability under ML decoding given that the m-th message is transmitted, denoted by $P_{e|m}$. The DS2 bound on the conditional error

probability (see [31], [32], [94] and [96]) gets the form

$$P_{\mathrm{e}|m} \leq \left(\sum_{\mathbf{y} \in \mathcal{Y}^n} G_n^m(\mathbf{y}) p_n(\mathbf{y}|\mathbf{x}_m) \right)^{1-\rho} \cdot \left\{ \sum_{m' \neq m} \sum_{\mathbf{y} \in \mathcal{Y}^n} G_n^m(\mathbf{y})^{1-\frac{1}{\rho}} p_n(\mathbf{y}|\mathbf{x}_m) \left(\frac{p_n(\mathbf{y}|\mathbf{x}_{m'})}{p_n(\mathbf{y}|\mathbf{x}_m)} \right)^{\lambda} \right\}^{\rho} \tag{2.3}$$

where $\mathcal{Y}$ is a discrete output-alphabet, $G_n^m(\mathbf{y})$ is an arbitrary non-negative function of $\mathbf{y} \in \mathcal{Y}^n$, and $0 \leq \rho \leq 1$ and $\lambda \geq 0$ are arbitrary real-valued parameters. Here $p_n(\mathbf{y}|\mathbf{x})$ designates the transition probability of the channel where $\mathbf{x} \in \mathcal{C}$ is the transmitted codeword and $\mathbf{y} \in \mathcal{Y}^n$ is the received vector. Notice that the bound in (2.3) holds for an arbitrary channel regardless of its input alphabet.

Consider now the class of memoryless symmetric channels with an input-alphabet $\mathcal{X}$. According to Proposition 2.1, $P_{\mathrm{e}|m}$ is independent of the transmitted message m. We further assume that $G_n^0(\mathbf{y})$ is expressed in the following product form:

$$G_n^0(\mathbf{y}) = \prod_{i=1}^{n} g(y_i)$$

where $g : \mathcal{Y} \to \mathbb{R}_+$ is an arbitrary non-negative function which is defined over the set $\mathcal{Y}$. The following bound on the decoding error probability is obtained for a discrete output alphabet (a similar proposition can be stated for channels with a continuous output alphabet):

Proposition 2.2 Consider an (n, k) linear block code $\mathcal{C}$ whose transmission takes place over a memoryless symmetric channel. Assume that the channel input and output alphabets are $\mathcal{X}$ and $\mathcal{Y}$, respectively, and let p be the transition probability of the channel. Then the block error probability of the code $\mathcal{C}$ under ML decoding, P_{e}, satisfies

$$P_{\mathrm{e}} \leq \left(\sum_{y \in \mathcal{Y}} g(y) p(y|0) \right)^{n(1-\rho)} \left\{ \sum_{m' \neq 0} \prod_{i=1}^{n} \sum_{y \in \mathcal{Y}} g(y)^{1-\frac{1}{\rho}} p(y|0)^{1-\lambda} p(y|x_{m',i})^{\lambda} \right\}^{\rho} \tag{2.4}$$

where $g : \mathcal{Y} \to \mathbb{R}_+$ is an arbitrary non-negative real function, $\lambda \geq 0$, and $0 \leq \rho \leq 1$ are arbitrary real-valued parameters.

Proof: See Appendix 2.C. ∎

2.2.2 Performance evaluation of ensembles of linear block codes

Definition 2.3 (Composition of a vector) Let $\mathbf{c}$ be a vector whose components are symbols in an alphabet $\mathcal{X}$ of size q. Let us assume without loss of generality that $\mathcal{X} = \{0, \ldots, q-1\}$. The composition of $\mathbf{c}$, denoted by $\mathbf{t} = \mathbf{t}(\mathbf{c})$, is a vector $\mathbf{t} = (t_0, t_1, \ldots, t_{q-1})$ where t_x (for $x \in \mathcal{X}$) counts the number of symbols in $\mathbf{c}$ that are equal to x.

Definition 2.4 (Complete composition spectrum) Let $\mathcal{C}$ be a linear block code of length n over an alphabet $\mathcal{X}$. The complete composition spectrum is the sequence $\{|\mathcal{C}_{\mathbf{t}}|\}$ where $|\mathcal{C}_{\mathbf{t}}|$ is the number of codewords whose composition is $\mathbf{t}$, and $\mathbf{t}$ ranges over the set $\mathcal{H}$ of all possible compositions over $\mathcal{X}^n$.

The existence of the all-zero codeword is clear. Consequently, the set $\mathcal{H}$ denotes in the followings the entire set of possible compositions except for the one of the all-zero codeword. The following lemma considers the error probability under ML decoding of an ensemble of linear block codes.

Lemma 2.3 Let $\mathcal{E}$ be an ensemble of linear block codes with block length n, and let $d_{\min}$ be the random variable designating the minimum Hamming distance of a randomly selected codebook $\mathcal{C}$ from this ensemble. Assume that there exist non-negative numbers D_n and ϵ_n, such that

$$\sum_{\{\mathbf{t} \in \mathcal{H}:\ n - t_0 \leq D_n\}} \mathsf{E}\big[|\mathcal{C}_{\mathbf{t}}|\big] \leq \epsilon_n \tag{2.5}$$

where $\mathsf{E}\big[\,|\mathcal{C}_{\mathbf{t}}|\,\big]$ denotes the expected number of codewords in $\mathcal{C}$ with composition $\mathbf{t}$, and $\mathcal{H}$ denotes the entire set of compositions except for the one of the all-zero codeword. Then, the block error probability under ML decoding satisfies

$$P_{\mathrm{e}} \leq \Pr(\text{ error } \mid d_{\min} > D_n) + \epsilon_n. \tag{2.6}$$

Proof:

$$\begin{aligned} P_{\mathrm{e}} &= \Pr(\text{ error } \mid d_{\min} > D_n)\Pr(d_{\min} > D_n) \\ &\quad + \Pr(\text{ error } \mid d_{\min} \leq D_n)\Pr(d_{\min} \leq D_n) \\ &\leq \Pr(\text{ error } \mid d_{\min} > D_n) + \Pr(d_{\min} \leq D_n). \end{aligned}$$

Let $\mathcal{C}$ be a codebook, chosen uniformly at random from the code ensemble $\mathcal{E}$, and let $w_{\mathrm{H}}(\mathbf{c})$ denote the Hamming weight of a codeword $\mathbf{c} \in \mathcal{C}$. Then, the union bound gives that

$$\begin{aligned}\Pr(d_{\min} \leq D_n) &\leq \sum_{\{\mathbf{c}\neq\mathbf{0}:\ w_{\mathrm{H}}(\mathbf{c})\leq D_n\}} \Pr(\mathbf{c} \in \mathcal{C}) \\ &= \sum_{\{\mathbf{t}\in\mathcal{H}:\ n-t_0\leq D_n\}} \ \sum_{\{\mathbf{c}:\ \mathbf{t}(\mathbf{c})=\mathbf{t}\}} \mathsf{E}\big[1_{\{\mathbf{c}\in\mathcal{C}\}}\big] \\ &= \sum_{\{\mathbf{t}\in\mathcal{H}:\ n-t_0\leq D_n\}} \mathsf{E}\big[|\mathcal{C}_{\mathbf{t}}|\big] \qquad (2.7)\end{aligned}$$

where $1_{\{\mathbf{c}\in\mathcal{C}\}}$ denotes the indicator of the event $\{\mathbf{c} \in \mathcal{C}\}$, and the last equality follows by converting the inner summation to an expectation. ■

Later in this section, we obtain upper bounds for the first term on the RHS of (2.6). These bounds are expressed in terms of the composition spectrum of the considered code ensemble, and they serve to find a suitable tradeoff between the parameters D_n and ϵ_n introduced in Lemma 2.3. More explicitly, since these two parameters are related, one wishes to increase the parameter D_n while maintaining small values of ϵ_n. The continuation to this section relies on Lemma 2.3 for the derivation of some bounds, and exemplify their use to regular LDPC code ensembles.

The following theorem provides an upper bound on the decoding error probability for ensembles of linear block codes whose transmission takes place over memoryless symmetric channels.

Theorem 2.1 Under the assumptions and notation in Proposition 2.2 and Lemma 2.3, the block error probability under ML decoding satisfies

$$\begin{aligned}P_{\mathrm{e}} \leq &\left(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\right)^{n(1-\rho)} \left(\sum_{\mathbf{t}\in\mathcal{H}:\ n-t_0>D_n} \mathsf{E}\Big[|\mathcal{C}_{\mathbf{t}}| \ \Big|\ d_{\min} > D_n\Big] \prod_{x\in\mathcal{X}} \big(s_{\lambda,\rho}(x)\big)^{t_x}\right)^{\rho} \\ &+ \epsilon_n \qquad (2.8)\end{aligned}$$

where

$$s_{\lambda,\rho}(x) \triangleq \sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}} p(y|0)^{1-\lambda} p(y|x)^{\lambda}, \quad x \in \mathcal{X} \qquad (2.9)$$

and $\mathsf{E}\big[\,|\mathcal{C}_{\mathbf{t}}| \ \big|\ d_{\min} > D_n\big]$ denotes the conditional expected number of codewords whose composition is equal to $\mathbf{t}$ (where the expectation is with respect to the choice of the codebook $\mathcal{C}$ from the ensemble $\mathcal{E}$) under the requirement that the minimal Hamming weight of the randomly selected codebook is larger than D_n.

Proof: From Proposition 2.2 and (2.9), we get the following upper bounding on the first summand in (2.6):

$$\begin{aligned}&\Pr(\text{ error } \mid d_{\min} > D_n)\\ &\quad\le \left(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\right)^{n(1-\rho)} \mathsf{E}\left[\left(\sum_{\mathbf{t}\in\mathcal{H}}\sum_{\mathbf{c}\in\mathcal{C}_\mathbf{t}}\prod_{i=1}^{n} s_{\lambda,\rho}(c_i)\right)^{\rho} \;\middle|\; d_{\min} > D_n\right]\end{aligned}$$

where $\mathcal{C}_\mathbf{t}$ is the set of all codewords in a codebook $\mathcal{C}$ whose composition is $\mathbf{t}$. Notice that the double summations on the RHS of the last inequality, over compositions $\mathbf{t}$ and codewords $\mathbf{c} \in \mathcal{C}_\mathbf{t}$, is equivalent to a single summation over all the non-zero codewords. Using Jensen's inequality, $\mathsf{E}[X^\rho] \le \big(\mathsf{E}[X]\big)^\rho$ for $0 \le \rho \le 1$, then

$$\begin{aligned}&\Pr(\text{ error } \mid d_{\min} > D_n)\\ &\quad\le \left(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\right)^{n(1-\rho)}\\ &\qquad\cdot\left(\sum_{\mathbf{t}\in\mathcal{H}} \mathsf{E}\left[\sum_{\mathbf{c}\in\mathcal{C}_\mathbf{t}}\prod_{x\in\mathcal{X}}\big(s_{\lambda,\rho}(x)\big)^{t_x} \;\middle|\; d_{\min} > D_n\right]\right)^{\rho}\\ &\quad= \left(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\right)^{n(1-\rho)}\\ &\qquad\cdot\left(\sum_{\mathbf{t}\in\mathcal{H}} \mathsf{E}\Big[\big|\mathcal{C}_\mathbf{t}\big| \;\Big|\; d_{\min} > D_n\Big]\prod_{x\in\mathcal{X}}\big(s_{\lambda,\rho}(x)\big)^{t_x}\right)^{\rho}. \qquad (2.10)\end{aligned}$$

For all codewords whose composition $\mathbf{t}$ satisfies $n - t_0 \le D_n$, their Hamming weight is not larger than D_n. Hence

$$\mathsf{E}\Big[\big|\mathcal{C}_\mathbf{t}\big| \;\Big|\; d_{\min} > D_n\Big] = 0, \quad \forall\, \mathbf{t} \in \mathcal{H} :\; n - t_0 \le D_n \qquad (2.11)$$

and the bound in (2.8) follows from Lemma 2.3, and (2.10) and (2.11). ■

The following theorem is a particularization of Theorem 2.1:

Theorem 2.2 Under the assumptions and notation in Proposition 2.2 and Lemma 2.3, the block error probability satisfies

$$P_\mathrm{e} \le\; q^{-nE_\mathrm{r}\left(R+\frac{\log_q \alpha_q(\mathcal{C},D_n)}{n}\right)} \;+\; \epsilon_n \qquad (2.12)$$

where n and R are the block length and code rate (measured in q-ary symbols per

channel use), respectively, and

$$E_{\mathrm{r}}(R) \triangleq \max_{0\leq\rho\leq 1} \left(E_0(\rho) - \rho R\right)$$

$$E_0(\rho) \triangleq -\log_q \left(\sum_{y\in\mathcal{Y}} \left(\frac{1}{q} \sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}} \right)^{1+\rho} \right)$$

$$\alpha_q(\mathcal{C}, D_n) \triangleq \max_{\{\mathbf{t}\in\mathcal{H}:\ n-t_0>D_n\}} \left\{ \frac{\mathsf{E}\left[|\mathcal{C}_{\mathbf{t}}| \ \middle|\ d_{\min} > D_n \right]}{q^{-n(1-R)}\binom{n}{\mathbf{t}}} \right\}. \tag{2.13}$$

Proof: See Appendix 2.D. ∎

A similar theorem can be stated for memoryless symmetric channels with continuous-output alphabets, where sums are replaced by integrals.

The bound in Theorem 2.2 is based on two summands. The first is an adaptation of the SFB to non-binary linear block codes which applies to the codebooks whose minimum distance exceeds an arbitrary threshold D_n. The second term relates to the probability that a randomly selected codebook from the ensemble has a minimum Hamming distance which does not exceed D_n. As a result, the second term on the RHS of (2.12) does not depend on the communication channel, but only on the code ensemble and the arbitrary threshold D_n. This partitioning differs from [11] and [78] where no such separation of codebooks is used. The SFB in [11] and [78] is combined with a union bound which corresponds to all pairwise error probabilities of relevant codewords and it depends on the communication channel. Following Example 2.2, the SFB in [11] can be considered as a particular case of Theorem 2.2 (the same goes for [37] where the considered modulo-additive noise channel is also symmetric according to Definition 2.1).

In general, the conditional expectation of the composition spectrum given that the minimum Hamming distance exceeds a certain positive threshold D_n (i.e., $\mathsf{E}\big[|\mathcal{C}_{\mathbf{t}}|\big|d_{\min} > D_n\big]$) is not available. Nevertheless, it is possible to use the inequality

$$\begin{aligned} \mathsf{E}\Big[|\mathcal{C}_{\mathbf{t}}|\Big] &\geq \mathsf{E}\Big[|\mathcal{C}_{\mathbf{t}}| \mid d_{\min} > D_n\Big]\ \Pr(d_{\min} > D_n) \\ &\geq \mathsf{E}\Big[|\mathcal{C}_{\mathbf{t}}| \mid d_{\min} > D_n\Big]\ (1-\epsilon_n). \end{aligned} \tag{2.14}$$

where the LHS of this inequality requires the knowledge of the expectation of the complete composition spectrum $\mathsf{E}\big[|\mathcal{C}_{\mathbf{t}}|\big]$. Applying (2.14) to the RHS of (2.8), gives a looser version of the bounds in Theorem 2.1 and 2.2 but is more amenable to analysis. The inequality in (2.14) is valid when expurgation of codebooks is considered. The expurgated ensemble is constructed by removing all codebooks whose minimum

Hamming distance is not larger than D_n. Since all the codebooks in the expurgated ensemble have a minimum distance greater than D_n, then the additive term ϵ_n on the RHS of (2.8) vanishes.

Consider an ensemble of linear block codes, and choose a codebook from this ensemble uniformly at random. We further assume that the probability that a vector is a codeword only depends on its Hamming weight (so all vectors of a fixed composition are codewords with equal probability). As a result, the expected complete composition spectrum $\mathsf{E}\,|\mathcal{C}_{\mathbf{t}}|$ satisfies

$$\mathsf{E}\Big[|\mathcal{C}_{\mathbf{t}}|\Big] = P(n-t_0)\binom{n}{\mathbf{t}} \tag{2.15}$$

where $P(l)$ denotes the probability that a word whose Hamming weight is l, forms a codeword in a randomly selected codebook from the ensemble. Assuming (2.15), the evaluation of α_q in Theorem 2.2 is considerably reduced.

In the following, we introduce an improvement over the bound in Theorem 2.2:

Theorem 2.3 Under the assumptions and notation in Proposition 2.2 and Lemma 2.3, for ensembles satisfying (2.15), the block error probability satisfies

$$P_{\mathrm{e}} \leq A(\rho)^{n(1-\rho)} \left(\sum_{D_n < l \leq n} \frac{P(l)}{1-\epsilon_n} \binom{n}{l} B(\rho)^{n-l} C(\rho)^l \right)^{\rho} + \epsilon_n \tag{2.16}$$

where $0 \leq \rho \leq 1$, ϵ_n is defined in (2.6), and

$$A(\rho) \triangleq \sum_{y \in \mathcal{Y}} \left(\frac{1}{q} \sum_{x \in \mathcal{X}} p(y|x)^{\frac{1}{1+\rho}} \right)^{1+\rho}$$

$$B(\rho) \triangleq \sum_{y \in \mathcal{Y}} \left(\frac{1}{q} \sum_{x \in \mathcal{X}} p(y|x)^{\frac{1}{1+\rho}} \right)^{\rho-1} \left(\frac{1}{q} \sum_{x \in \mathcal{X}} p(y|x)^{\frac{2}{1+\rho}} \right)$$

$$C(\rho) \triangleq qA(\rho) - B(\rho).$$

Proof: See Appendix 2.E. ∎

Remark 2.3 For the particular case of binary linear block codes, the bound provided in Theorem 2.3 does not require the symmetry assumption on the considered ensemble in (2.15). For this case, the same derivation holds while setting

$$P(l) \triangleq \frac{\mathsf{E}\big[\,|\mathcal{C}_l|\,\big]}{\binom{n}{l}}, \;\; D_n < l \leq n$$

where $\mathsf{E}\big[\,|\mathcal{C}_l|\,\big]$ denotes the expected number of codewords whose Hamming weight is l.

2.2.3 Performance of non-binary regular LDPC ensembles

The non-binary (c, d)-regular LDPC code ensemble, proposed by Gallager in [44, Ch. 5], is considered with the q-ary symmetric channel and the AWGN channel with a q-ary PSK modulation (both channels are symmetric according to Definition 2.1). The Gallager ensemble is defined using a sparse parity-check matrix with *binary* elements. This matrix is regular, having c ones in each column and d ones in each row. The LDPC ensemble is constructed as follows:

1. Divide the parity check matrix into c consecutive sub-matrices. All the sub-matrices have n columns and $\frac{cn}{d}$ rows.
2. Fill the first sub-matrix with ones in a descending order.
3. All other sub-matrices are chosen as random permutations of the first sub-matrix.
4. Parity-check equations are evaluated using a modulo-q arithmetics.

The following lemma is provided in [44] which implies an upper bound on the complete composition spectrum satisfying the condition in (2.15):

Lemma 2.4 Consider the regular non-binary LDPC ensemble of Gallager. Let $\mathbf{x}$ be a vector of weight $l > 0$. The probability $P(l)$ that the vector $\mathbf{x}$ is a codeword of a codebook which is selected uniformly at random from the ensemble, is upper bounded by

$$P(l) \leq \left(\frac{\exp\left(\frac{n}{d}\left(\mu_q(s) - s\mu_q'(s) + (d-1)\ln q\right)\right)}{\binom{n}{l}(q-1)^l} \right)^c \tag{2.17}$$

where

$$\mu_q(s) \triangleq \ln\left(\frac{\left(1 + (q-1)e^s\right)^d + (q-1)\left(1 - e^s\right)^d}{q^d} \right)$$

and s is a real number given by the solution of the following equation

$$\frac{n}{d}\mu_q'(s) = l. \tag{2.18}$$

Note, that the bound in (2.17) is valid for all s, not only for the one satisfying (2.18) which yields the minimum bound in (2.17). Using the change of variables $s = \ln\frac{1-u}{1+(q-1)u}$, $-\frac{1}{q-1} \leq u \leq 1$, in (2.18), results in the following polynomial equation:

$$\left(\frac{wq}{n} - 1\right)u^d + u^{d-1} + u + \frac{wq}{n(q-1)} - 1 = 0.$$

For $q > 2$, this equation has a single root in the interval $\left[-\frac{1}{q-1}, 1\right]$ (the details concerning the evaluation of the RHS of (2.17) for the binary case are provided in [93]).

In the following, we obtain the exact composition spectrum of the regular LDPC code ensembles of Gallager. This derivation serves to improve the tightness of the bounds on the error probability. The provided analysis generalizes [105] to non-binary codes. The exact enumeration for the binary case is already available in [18], as an intermediate result, although its main interest is in asymptotic analysis (This analysis can be traced even to Gallager [45]).

Lemma 2.5 Under the assumptions and notation in Lemma 2.4, the probability $P(l)$ satisfies

$$P(l) = \left(\frac{A_l}{\binom{n}{l}(q-1)^l} \right)^c, \quad 2 \leq l \leq n \tag{2.19}$$

where

$$\sum_{2 \leq l \leq n} A_l X^l \triangleq \big(A^*(X)\big)^{\frac{n}{d}} \tag{2.20}$$

$$A^*(X) \triangleq 1 + \frac{1}{q} \sum_{l=2}^{d} \Big((q-1)^l + (q-1)(-1)^l \Big) \binom{d}{l} X^l. \tag{2.21}$$

Proof: See Appendix 2.F. ■

As suggested in [105], the numerical evaluation of the exponent in (2.20) is carried out, in all the examples studied in this chapter, via the binary method (see [61, p. 441]). This method makes the evaluation of the high-order powers of a polynomial relatively easy to compute.

The 1961 Gallager-Fano bound (see [94, 44]) and Lemma 2.4 imply an exponential bound (in terms of the block length) on the decoding error probability for the expurgated LDPC code ensemble. This expurgation removes all the codebooks whose minimal Hamming distance is below a certain threshold which scales linearly with the block length. This result is elaborated for the binary case by Miller and Burshtein [78]).

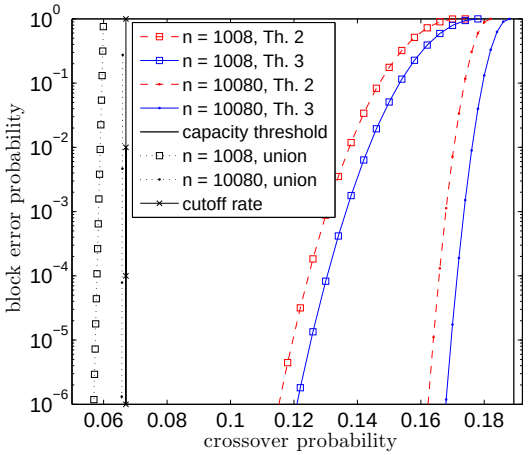
The following examples consider the Gallager ensembles of non-binary and (8, 16) regular LDPC codes where these ensembles are expurgated by removing all the codebooks whose minimum distance is not greater than a certain parameter D_n. The examples study upper bounds on the decoding error probability of these expurgated ensembles via the use of the upper bounds in Theorems 2.2 and 2.3. The exact composition spectrum of the non-expurgated LDPC code ensemble is evaluated via

Lemma 2.5, and then upper bounds on the composition spectrum of the expurgated ensembles are calculated via (2.14).

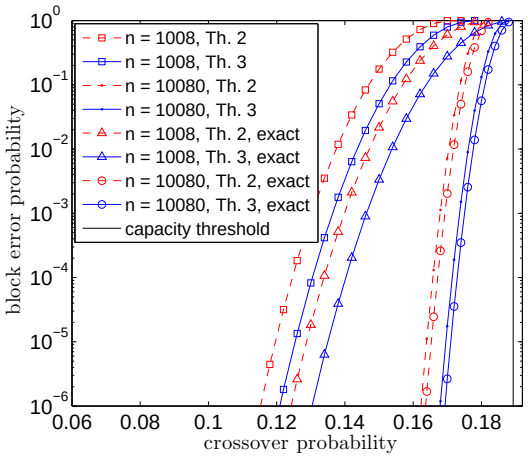
Example 2.4 (q-ary symmetric channels) Bounds on the block error probability for some expurgated LDPC code ensembles are presented in Figure 2.1 when the transmission takes place over a q-ary symmetric channel and ML decoding is performed. The performance bounds introduced in this chapter are compared with the union bound, and we also exemplify the uselessness of the union bound beyond the crossover probability which corresponds to the cutoff rate. More specifically, for a q-ary symmetric channel, the cutoff rate is given by

$$R_0 = 1 - 2\log_q\left(\sqrt{1-p} + \sqrt{p(q-1)}\right)$$

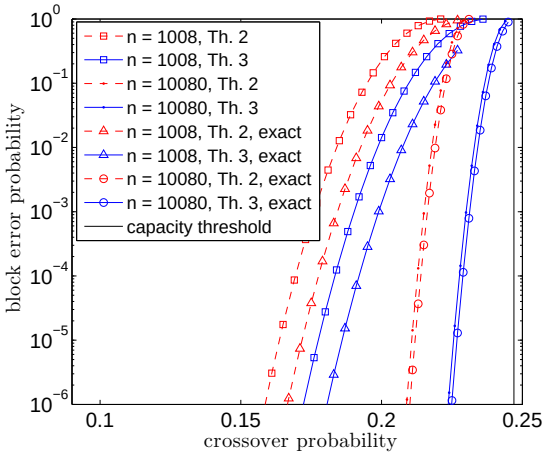
so the crossover probability which follows by setting the value of R_0 to the code rate (which is one-half symbol per channel use in Figure 2.1) is equal to $p = 0.0670$ and $p = 0.0739$ for quaternary and octal input alphabets, respectively. The union bound shown in the upper plot of Figure 2.1 (see plot (a)) has a sharp decline around the crossover probability which corresponds to the cutoff rate of the q-ary symmetric channel (i.e., around $p = 0.0670$ for $q = 4$). Plot (a) also exemplifies the potential application of the proposed bounds to assess the performance of efficient code ensembles which perform reliably at rates exceeding the cutoff rate of the channel. Figure 2.1(b) is focused on the improved bounds in Theorems 2.2 and 2.3, applied to the Gallager $(8, 16)$ regular and expurgated LDPC code ensemble with a quaternary alphabet and block lengths of $n = 1008$ and 10080 symbols. The ensemble spectrum is upper bounded via Lemma 2.4, and in addition it is exactly evaluated using Lemma 2.5; both options are applied in this example so that the improvement provided by the exact calculation of the composition spectrum is exemplified in this figure. The various choices of the parameter D_n and the resulting ϵ_n, which serves as an upper bound on the fraction of codebooks whose minimum distance is not larger than D_n, are detailed in Table 2.1(a). Since Theorem 2.3 is tighter than Theorem 2.2, then the minimal value of D_n for which Theorem 2.2 is useful is larger than the corresponding value which is calculated in conjunction with Theorem 2.3. Moreover, the considered bounds are further improved when the upper bound for the composition spectrum in Lemma 2.4 is replaced with the exact calculation in Lemma 2.5. The inferiority of the SFB in (2.12) is further pronounced for higher alphabets, as exemplified for octal signaling in Figure 2.1(c) (where the details with regard to the choices of D_n and ϵ_n values are given in Table 2.1(b)).



(a) $q = 4$



(b) $q = 4$



(c) $q = 8$

Figure 2.1: Upper bounds on the block error probability of the Gallager $(8, 16)$ regular and non-binary LDPC code ensembles with quaternary and octal input alphabets. The transmission takes place over a q-ary symmetric channel where $q = 4$ in plots (a) & (b) and $q = 8$ in plot (c). This figure refers to expurgated ensembles whose block lengths are 1008 and $10,080$ symbols.

Table 2.1: Parameters for Example 2.4

(a) Quaternary alphabet ($q = 4$).

Performance bound	Block length n (symbols)	D_n	ϵ_n (Lemma 2.4)	ϵ_n (Lemma 2.5)
Theorem 2.2	1008	173	0.1	10^{-11}
Theorem 2.3	1008	99	10^{-4}	10^{-11}
Theorem 2.2	10008	1834	0.11	10^{-17}
Theorem 2.3	10008	600	10^{-7}	10^{-17}

(b) Octal alphabet ($q = 8$).

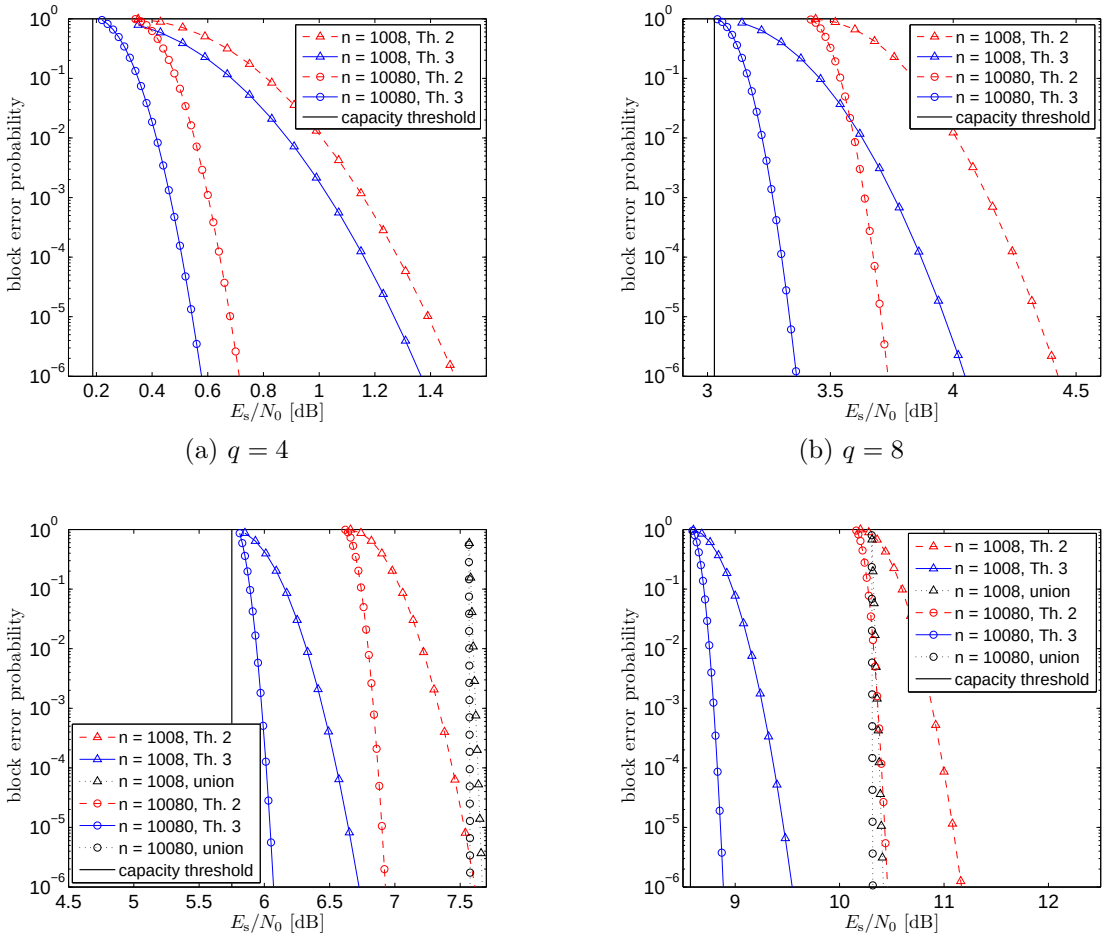
Performance bound	Block length n (symbols)	D_n	ϵ_n (Lemma 2.4)	ϵ_n (Lemma 2.5)
Theorem 2.2	1008	191	10^{-5}	10^{-14}
Theorem 2.3	1008	119	10^{-5}	10^{-14}
Theorem 2.2	10080	1951	10^{-9}	10^{-20}
Theorem 2.3	10080	887	10^{-9}	10^{-20}

Example 2.5 (AWGN channels with a q-ary PSK modulation) Upper bounds on the block error probability for for some expurgated LDPC code ensembles are depicted in Figure 2.2 when the transmission takes place over the AWGN channel with a q-ary PSK modulation. The alphabet size of these code ensembles is $q = 4, 8, 16$, and 32, and the examined parameters D_n of the expurgation are given in Table 2.2. It is evident that the SFB in Theorem 2.2 deteriorates as compared to the bound in Theorem 2.3. This deterioration is more dominant by increasing the alphabet size q. It is interesting to compare the studied bounds to the union bound which, for large block lengths, diverges at the cutoff rate of the communication channel. For alphabet cardinalities of $q = 4$ and $q = 8$, the cutoff rate corresponds to $\frac{E_s}{N_0}$ ratios of 2.46 dB and 5.05 dB, respectively, which exemplify the superiority of both derivations over the union bound. However, for alphabet cardinalities of $q = 16$ and $q = 32$, the SFB deteriorates considerably comparing to the bound provided in Theorem 2.3 and to the union bound which is depicted in Figure 2.2 and (d) (the SNR values which correspond to the cutoff rate for $q = 16$ and 32 are equal to 7.57 dB and 10.31 dB, respectively).

The reason for the deterioration of the SFB for large values of q is explained when looking into the rate term $\frac{1}{n} \log_q \alpha(\mathcal{C}, D_n)$. This term corresponds to the difference between the spectrum of the considered ensemble and the multinomial spectrum of the fully random code ensemble. This difference between the two composition spectra is depicted in Figure 2.3 as a function of $\frac{D_n}{n}$ for alphabet sizes of $q = 4$, 8, 16, and 32, and for block lengths of $n = 512$, 1008, and 10080 symbols. From Figure 2.3, this term is more pronounced by increasing the value of q. On the other hand, the bound in Theorem 2.3 does not exhibit such deterioration.

Table 2.2: D_n values for Example 2.5

Performance bound	Block length n (symbols)	D_n (q=4)	D_n (q=8)	D_n (q=16)	D_n (q=32)
Theorem 2.2	1008	186	191	191	191
Theorem 2.3	1008	38	34	15	12
Theorem 2.2	10080	1851	1951	1951	1951
Theorem 2.3	10080	282	216	132	102



(a) $q = 4$ (b) $q = 8$

(c) $q = 16$ (d) $q = 32$

Figure 2.2: Upper bounds on the block error probability under ML decoding of the $(8, 16)$-regular LDPC ensembles of Gallager with alphabet size of $q = 4$, 8, 16, and 32, whose transmission takes place over an AWGN channel with a q-ary PSK modulation. This figure depicts the upper bounds on the block error probability for the expurgated ensemble with block lengths of 1008 and $10,080$ symbols.

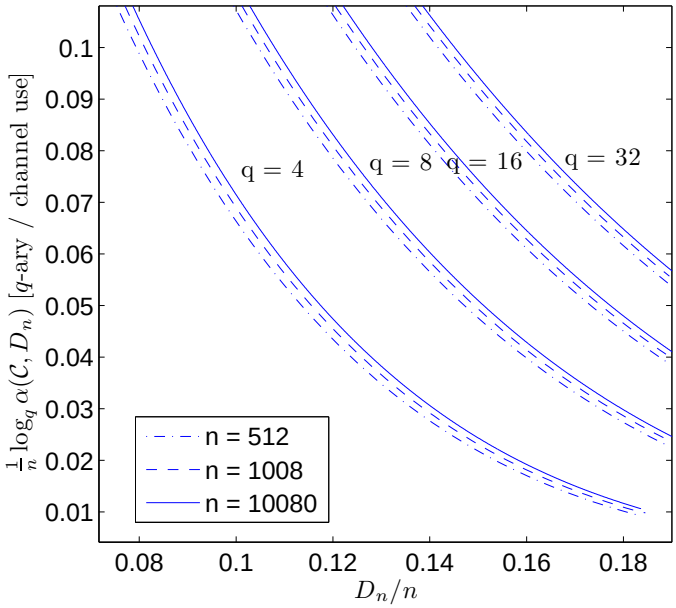


Figure 2.3: The term $\frac{1}{n}\log_q \alpha_q(\mathcal{C}, D_n)$ in (2.12) for the regular (8,16) LDPC ensemble of Gallager [44], depicted for alphabet sizes of $q = 4$, 8, 16, and 32, and block lengths of $n = 512$, 1008, and 10080 symbols.

Remark 2.4 Divsalar's bound [26, 27] is widely used when assessing the error performance of binary turbo-like code ensembles over the binary-input AWGN channel (see [94, Chapter 3.2.4] and references therein). This is due to the fact that the bound is given in a closed form, and its calculation does not involve any numerical integrations and parameter optimizations. The basic concept the bound is based on is the following:

$$\Pr(\text{error}) \leq \Pr(\text{error}, \mathbf{y} \in \mathcal{R}) + \Pr(\mathbf{y} \notin \mathcal{R})$$

where $\mathbf{y}$ is the received vector, and the region $\mathcal{R}$ is the n-dimensional sphere which is centered at a point along the line connecting the origin to the all-zero codeword, and whose radius is optimized analytically in order to get the tightest bound within its form. This technique was generalized by the authors to the non-binary setup by examining various regions in the complex observation space. In contrast to the binary case, not all the parameters could be optimized analytically. Moreover, the resulting bounds were not satisfactory as compared to the bounds presented in Example 2.5, and are therefore omitted.

Example 2.6 (A Comparison to lower bounds on the decoding error probability) The upper bound in Theorem 2.3 is compared in Figure 2.4 to the SP59 lower bound of Shannon [99], and the ISP lower bound in [115]. The regular LDPC code ensembles of Gallager are considered with octal alphabet cardinality and block lengths

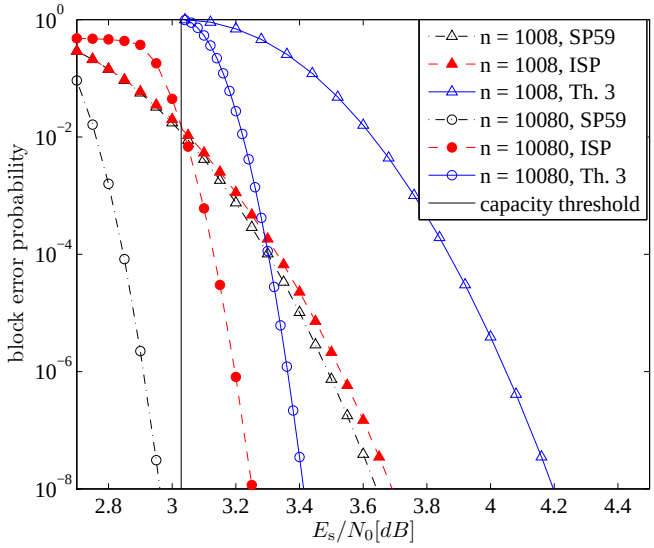
of 1008 and 10080 symbols, and the performance is studied over the AWGN channel with an 8-ary PSK modulation. In Figure 2.4(a), the upper bound in Theorem 2.3 is depicted for the Gallager (8,16) regular and expurgated LDPC code ensemble with octal alphabet (the bound is evaluated with the same parameters as in Table 2.2). In addition, the ultimate performance of a rate 0.5 code is assessed via the SP59 and the ISP lower bounds on the decoding error probability. For a block length of 1008 symbols, a negligible difference exists between the two considered lower bounds, and both of these bounds are about 0.5 dB away from the upper bound in Theorem 2.3 for all range of interest. For the larger block length of 10080 symbols, the gain of the ISP bound is about 0.25 dB as compared to the SP59 bound, and it is about 0.2 dB away from the upper bound (see Figure 2.4(a)). The comparison between the upper and lower bounds is further studied in Figure 2.4(b) for the Gallager (8,32) regular and expurgated LDPC code ensembles with block lengths of 1024 and 10080 symbols and octal alphabet. The design rate for these ensembles is 0.75 symbols per channel use. The upper bound in Theorem 2.3 is depicted with $D_n = 25$ and 95, respective to the studied block lengths. The ISP bound maintains its close proximity with the upper bound. The SP59 bound on the other hand deteriorates considerably for this case, and it is less informative than the capacity limit for both considered block lengths (see Figure 2.4(b)).

2.3 Gallager-type bounds for fully-interleaved fading channels with prefect CSI at the receiver

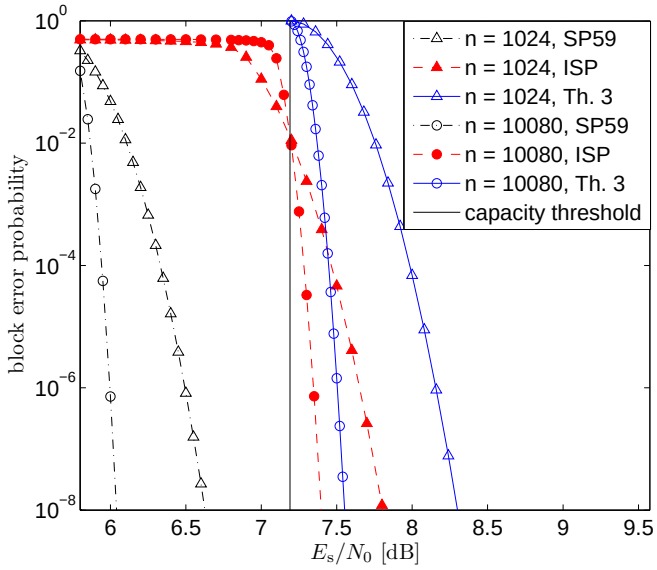
In the section, the error probability of a linear block code $\mathcal{C}$ is considered under ML decoding when transmission takes place over a fully-interleaved fading channel and perfect CSI is available at the receiver. The fading is assumed to be a continuous random variable (a similar framework is possible for the discrete case). Let $\mathcal{A}$ denote the set of possible fading samples, and $p(\mathbf{y}, \mathbf{a}|\mathbf{x})$ denote the conditional joint pdf of the received sequence $\mathbf{y} = (y_1, \ldots, y_n) \in \mathcal{Y}^n$ and the fading samples $\mathbf{a} = (a_1, \ldots, a_n) \in \mathcal{A}^n$ given that the transmitted codeword is $\mathbf{x} \in \mathcal{C}$. Due to an ideal symbol interleaving, the channel is memoryless and accordingly

$$p(\mathbf{y}, \mathbf{a}|\mathbf{x}) = \prod_{i=1}^{n} p(y_i|x_i, a_i)p(a_i)$$

where $p(y|x, a)$ is the single-letter conditional pdf of the channel, and $p(a)$ is the pdf of a fading sample. The following definition of symmetry is a generalization to the



(a) $R = 0.5$



(b) $R = 0.75$

Figure 2.4: A Comparison between the upper bound in Theorem 2.3 and the SP59 and ISP lower bounds on the decoding error probability for octal alphabet block codes whose transmission takes place over an AWGN channel with 8-ary PSK modulation. This figure depicts the upper and lower bounds on the block error probability for block lengths of 1008 and 10,080 symbols. The upper bounds are provided for expurgated $(8, 16)$ and $(8, 32)$ regular LDPC code ensembles.

one presented in Definition 2.1. This generalization is obtained by directly applying Definition 2.1 to a channel whose observations are the pair of the considered channel output and the fading sample.

Definition 2.5 Consider the fully-interleaved fading channel with an input-alphabet $\mathcal{X}$, and perfect CSI at the receiver. The channel, which is characterized by a transition pdf p, is symmetric if for every $a \in \mathcal{A}$, there exists a function $\mathcal{T}_a : \mathcal{Y} \times \mathcal{X} \to \mathcal{Y}$ which satisfies the following properties:

1. For every $x \in \mathcal{X}$, the function $\mathcal{T}_a(\cdot, x) : \mathcal{Y} \to \mathcal{Y}$ is bijective and with a Jacobian 1.

2. For every $x_1, x_2 \in \mathcal{X}$, the following equality holds:

$$p(y|x_1, a) = p(\mathcal{T}_a(y, x_2 - x_1)|x_2, a). \tag{2.22}$$

Notice that this definition of symmetry is a weaker notion compared to a one where there exists a function $\mathcal{T} : \mathcal{Y} \times \mathcal{X} \to \mathcal{Y}$ meeting the condition in (2.22) for every fading sample $a \in \mathcal{A}$. Nevertheless, this weaker notion is sufficient in order to prove that for the case at hand, the ML decoding error probability does not depend on the actual transmitted message. This is clearly expected since Definition 2.5 is a direct application of Definition 2.1 for the case at hand. The conditional decoding error probability for the m-th message under ML decoding as is given by

$$P_{\mathrm{e}|m} = \int_{\mathbf{a}} \int_{\mathbf{y} \in \Lambda_m^c(\mathbf{a})} p(\mathbf{y}, \mathbf{a}|\mathbf{x}_m)\, d\mathbf{y}\, d\mathbf{a} = \int_{\mathbf{a}} p(\mathbf{a}) \int_{\mathbf{y} \in \Lambda_m^c(\mathbf{a})} p(\mathbf{y}|\mathbf{x}_m, \mathbf{a})\, d\mathbf{y}\, d\mathbf{a} \tag{2.23}$$

where $\Lambda_m(\mathbf{a}) \subseteq \mathcal{Y}^n$ is the decision region under ML decoding given that the sequence of fading samples is $\mathbf{a} \in \mathcal{A}^n$. The proof of the independence of the decoding error probability on the transmitted codeword follows by showing that the inner integral in (2.23) is independent of the transmitted message m (this is accomplished for every sequence of fading sample sequence $\mathbf{a}$ in the same way as of the proof in Appendix 2.B).

Theorem 2.4 Under the assumptions and notation in Lemma 2.3, consider the case where transmission takes place over a symmetric, fully-interleaved fading channel with perfect CSI at the receiver. Let the channel input and output alphabets be $\mathcal{X}$ and $\mathcal{Y}$, respectively, and let p be the transition pdf of the channel. Then, the block

error probability under ML decoding satisfies

$$P_{\mathrm{e}} \leq \sum_{j=1}^{J} \Bigg(\sum_{\mathbf{t}\in\mathcal{H}_j:\ n-t_0>D_n} \mathsf{E}\Big[|\mathcal{C}_{\mathbf{t}}| \ \Big|\ d_{\min} > D_n\Big] \prod_{x\in\mathcal{X}} \left(\iint \psi_j(y,a)^{1-\frac{1}{\rho_j}} p(y,a|0)^{\frac{1-\lambda_j\rho_j}{\rho_j}} p(y,a|x)^{\lambda_j}\, dy\, da \right)^{t_x} \Bigg)^{\rho_j} + \epsilon_n \tag{2.24}$$

where $\{\mathcal{H}_j\}_{j=1}^{J}$ with an arbitrary $J \geq 1$ forms a partition of the set of compositions (except for the one which corresponds to the all-zero codeword) to J subsets, $\mathsf{E}\Big[|\mathcal{C}_{\mathbf{t}}| \ \Big|\ d_{\min} > D_n\Big]$ denotes the expectation of the complete composition spectrum under the assumption that $d_{\min} > D_n$, the functions $\psi_j : \mathcal{Y}\times\mathcal{A} \to \mathbb{R}$ are arbitrary non-negative tilting probability measures, and $0 \leq \rho_j \leq 1$ and $\lambda_j \geq 0$.

Proof: See Appendix 2.G. ■

Consider an ensemble which satisfies the symmetry property in (2.15), and let us choose $J = n$ and $\mathcal{H}_j = \{\mathbf{t} :\ n - t_0 = j\}$. By using calculus of variations, the optimum tilting measures ψ_j for $D_n < j \leq n$, are given by

$$\psi_j(y,a) = \alpha_{j,0}\, p(y,a|0) \left(1 + \sum_{x\in\mathcal{X}_*} \alpha_{j,x} \left(\frac{p(y,a|x)}{p(y,a|0)} \right)^{\lambda_j} \right)^{\rho_j}, \quad \lambda_j \geq 0,\ 0 \leq \rho_j \leq 1$$

where the parameters $\alpha_{j,x}$, $x \in \mathcal{X}^*$ are given by

$$\alpha_{j,x} \triangleq \frac{\frac{j}{n} \iint \psi_j(y,a)^{1-\frac{1}{\rho_j}} p(y,a|0)^{\frac{1}{\rho_j}}\, dy\, da}{(1-\frac{j}{n}) \sum_{x\in\mathcal{X}^*} \iint \psi_j(y,a)^{1-\frac{1}{\rho_j}} p(y,a|0)^{\frac{1-\lambda_j\rho_j}{\rho_j}} p(y,a|x)^{\lambda_j}\, dy\, da}$$

and $\alpha_{j,0}$ are determined such that ψ_j are probability measures. The numerical evaluations of such bounds result in a tedious numerical process. It is therefore of interest to seek for probability tilting measures for which the integration in (2.24) has a closed form expression. Exponential upper bounds on the ML decoding error probability of binary linear block codes that operate over the binary-input fully-interleaved Rician fading channel with perfect CSI at the receiver were derived in [58]. These bounds are reasonably tight in a certain portion of the rate region exceeding the cutoff rate, and do not require numerical integrations involved in the evaluation of the optimal DS2-based bound. In the following example, the technique in [58] is generalized and applied to non-binary linear block codes whose transmission takes place over a fully-interleaved Rician fading channel with a q-ary PSK modulation.

Example 2.7 (A fully-interleaved Rician fading channel with PSK modulation) Consider the class of fully-interleaved Rician fading channels with an additive white Gaussian noise. A codeword $\mathbf{x} = (x_1, \ldots, x_n)$ with a block length n and codeword symbols over the alphabet $\mathcal{X} = \{0, 1, \ldots, q-1\}$ is transmitted over a discrete-time memoryless channel. The received sequence $\mathbf{y} = (y_1, \ldots, y_n) \in \mathbb{C}^n$ satisfies

$$y_k = A_k \sqrt{\frac{2E_{\mathrm{s}}}{N_0}} \exp\left(\frac{2\pi i}{q} x_k\right) + N_k, \quad k = 1, \ldots, n. \tag{2.25}$$

Here A_k is a Rician random variable with a parameter K, and $N_k = N_k^{\mathrm{r}} + jN_k^{\mathrm{i}}$, where N_k^{r} and N_k^{i} are statistically independent Gaussian random variables with a zero mean and a unit variance. The non-negative real-valued parameter K designates the power ratio between the direct and the diffused paths, $N_0/2$ is the two sided power density spectrum of the additive white Gaussian noise, and E_{s} is the energy per transmitted coded symbol. The symmetry of the considered channel is guaranteed by the q-ary PSK modulation and the AWGN noise. Following [58], a sub-optimal DS2 bound is suggested for the case at hand. To this end, the exponential tilting measure

$$\psi_j(y, a) = \frac{\frac{\alpha_j}{2\pi} \exp\left(-\frac{\alpha_j}{2}\left|y - a u_j \sqrt{\frac{2E_{\mathrm{s}}}{N_0}}\right|^2 - \frac{\alpha v_j^2 a^2 E_{\mathrm{s}}}{N_0}\right) p(a)}{\int_0^\infty p(a) \exp\left(-\frac{\alpha v_j^2 a^2 E_{\mathrm{s}}}{N_0}\right) da}, \quad y \in \mathbb{C},\ a \geq 0 \tag{2.26}$$

where, for $1 \leq j \leq J$, v_j and α_j are non-negative real-valued parameters, and u_j is a complex-valued parameter. Substituting the exponential tilting measure ψ_j into (2.24) provides an upper bound on the error probability which is expressed in a closed form (see Appendix 2.H). The performance of the (8,16) regular non-binary LDPC ensemble of Gallager [44] with block lengths of $n = 1008$ and $n = 10080$ symbols is provided in Figure 2.5 using the bound in Theorem 2.4, in addition to the union bound. The bound in (2.24) is evaluated with $J = 6$ and the partitioning of the set of compositions is done according to their Hamming weights where the boundaries of this partitioning are set to Hamming weights of 350, 425, 500, 575, and 600 for a block length of 1008 symbols (the corresponding boundaries for a block length of 10080 symbols are set to 3500, 4250, 5000, 5750, and 6000). The performance bounds refer to a quaternary input-alphabet $q = 4$ and a fully-interleaved Rayleigh fading channel (see Figure 2.5(a)), and for octal input-alphabet $q = 8$ and a Rician fading channel with $K = 2$ (see Figure 2.5(b)). In both plots the non-expurgated ensemble is considered, while in plot (a) the performance for an expurgated ensemble with $D_n = 100$ (with a corresponding $\epsilon_n = 10^{-5}$ in Theorem 2.4) is also presented for a block length of 1008 symbols. In both plots, the union bound diverges bellow the cutoff rate which corresponds to E_{s}/N_0 thresholds of 5.1 dB and 7.18 dB respectively (the capacity

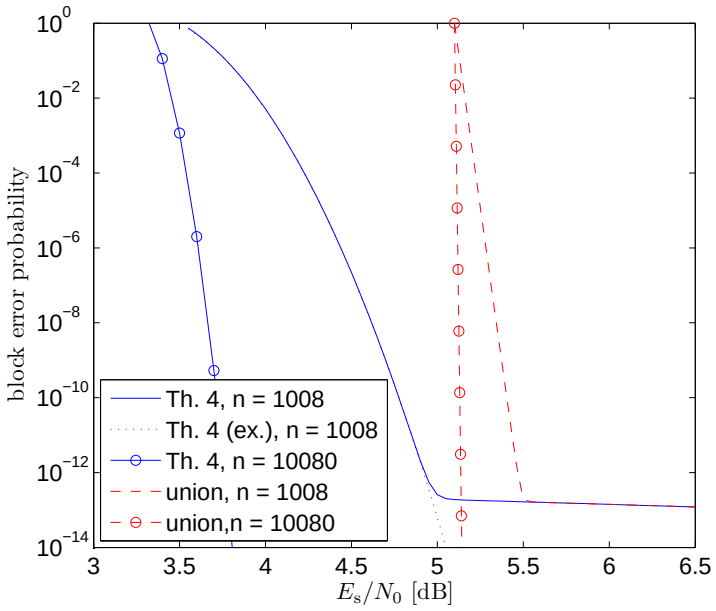
corresponds to thresholds of 1.86 dB and 4.21 dB, respectively). Although the bound in Theorem 2.4 is not informative (for the considered example) up to the ultimate channel capacity, it is for a block length of 1008 symbols 0.9 dB and 1 dB better than the union bound in Figure 2.5(a) and 1.2 dB and 1.3 dB in Figure 2.5(b) at block error probabilities of 10^{-6}, and 10^{-4}, respectively (for a block length of 10080 symbols the bound in Theorem 2.4 is better than the union bound by 1.5 dB and 1.8 dB, for quaternary and octal alphabets, respectively, at the considered block error probabilities).

Example 2.8 (A fully-interleaved Rayleigh fading channel with PSK modulation and maximal ratio combining) Consider the class of fully-interleaved Rayleigh fading channels with maximal ratio combining (MRC) space diversity of order L. The receiver sequence is as in (2.25) where the fading samples, A_k, are distributed according to the following pdf:

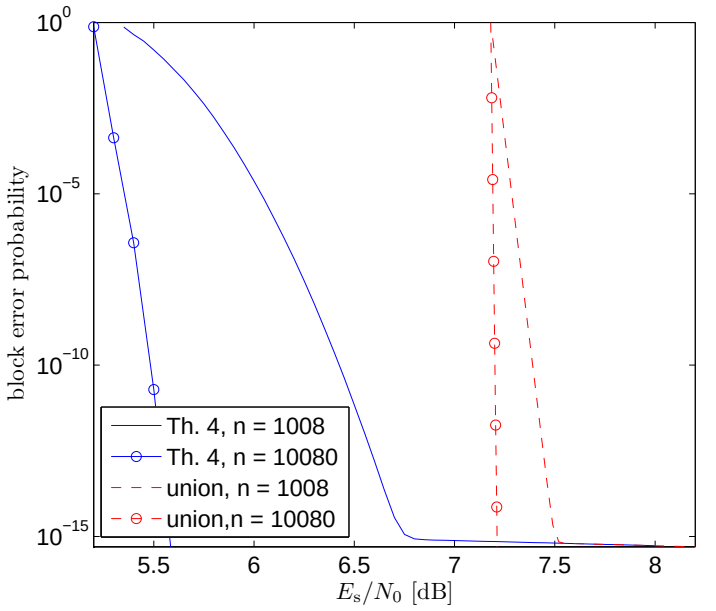
$$p(a) = \frac{2L^L a^{2L-1} \exp\left(-La^2\right)}{(L-1)!}, \ a \geq 0. \tag{2.27}$$

Note that $\frac{E_s}{N_0}$ in (2.25) refers to the stage after the MRC module. A closed-form expression for the upper bound on the block error rate, based on Theorem 2.4 and an exponential tilting measure is suggested (see Appendix 2.I). Consider the (8,16) regular and non-binary LDPC code ensemble of Gallager [44] with octal alphabet and a block length of 1008 symbols. Upper bounds on the decoding error probability of this ensemble with various diversity orders L are shown in Figure 2.6. The bound provided in Theorem 2.4 is compared with the union bound for MRC diversity with $L = 1$ to 4 antennas. Both bounds coincide in the error floor region which is considerably low for the considered ensemble. The union bound is informative only below the cutoff rate, which corresponds to E_s/N_0 of 8.51, 6.76, 6.18, and 5.90 dB for $L = 1, 2, 3$ and 4 receiving antennas. The bound provided in Theorem 2.4 is not informative up to the ultimate channel capacity (which corresponds to E_s/N_0 of 4.94, 4.00, 3.68, and 3.30 dB, respectively). Nevertheless, the bound in Theorem 2.4 outperforms the union bound by 1.33 dB at a block error rate of 10^{-4} when there is a single antenna at the receiver, and by 1.02 dB when $L = 4$ receiving antennas are used.

Example 2.9 (A comparison of upper and lower bounds) The DS2 upper bound in Theorem 2.4 is compared in this example to an improved sphere-packing (ISP) lower bound on the ultimate error performance of finite-length codes (see [115]). The bounds are compared for block codes whose transmission takes place over the fully



(a) $q = 4, K = 0$



(b) $q = 8, K = 2$

Figure 2.5: Upper bounds on the block error probability under ML decoding for the $(8, 16)$-regular LDPC ensemble of Gallager, whose transmission takes place over a fully-interleaved Rician fading channel with q-ary PSK modulation and perfect CSI at the receiver. Both plots refer to the non-expurgated ensemble, and the performance of an expurgated ensemble with $D_n = 100$ is also presented in plot (a) for comparison.

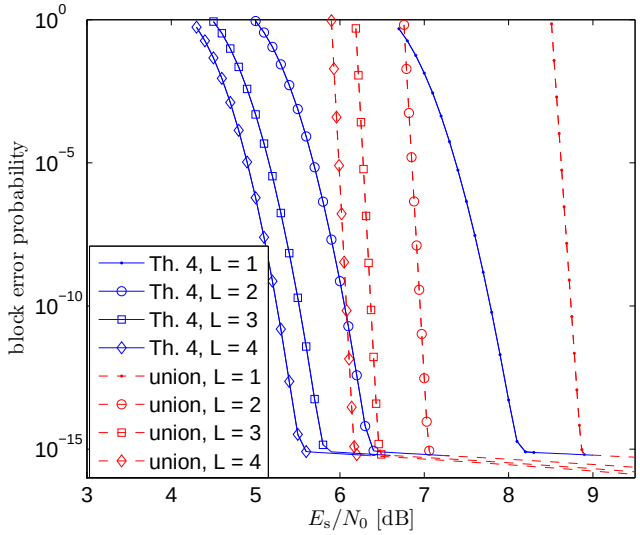


Figure 2.6: Upper bounds on the block error probability under ML decoding for the $(8, 16)$-regular LDPC ensemble of Gallager with octal alphabet and a block length of 1008 symbols. The transmission takes place over a fully-interleaved Rayleigh fading channel with 8-ary PSK modulation, perfect CSI and maximal ratio combining (MRC) at the receiver. The figure depicts the performance for MRC diversity with $L = 1$ to $L = 4$ antennas at the receiver.

interleaved Rayleigh fading channels with a quadrature-phase shift-keying (QPSK) modulation and perfect CSI at the receiver. The DS2 bound is evaluated with the sub-optimal exponential tilting measure in (2.26) for the (8,16) regular LDPC code ensembles of Gallager with block lengths of 1008 and 10080 symbols. The bounds are plotted in Figure 2.7 jointly with union bounds as a reference. The ultimate error performance using a rate–0.5 code with the considered block lengths is evaluated using the ISP lower bound [115]. For the two block lengths considered in this example, the ISP bound is more informative than the capacity threshold for decoding error probabilities below 10^{-2}. For a block length of 1008 symbols, the gap between the ISP lower bound and the sub-optimal DS2 upper bound is about 2.0 dB for a block error rate of 10^{-4}. For a block length of 10080 symbols, this gap is reduced to about 1.5 dB. Note that the use of the upper bound in Theorem 2.4 closes the 3 dB gap between the union upper bound and the respective ISP lower bound to only 1.5 dB while referring to a block length of 10080 symbols and a block error probability of 10^{-4}.

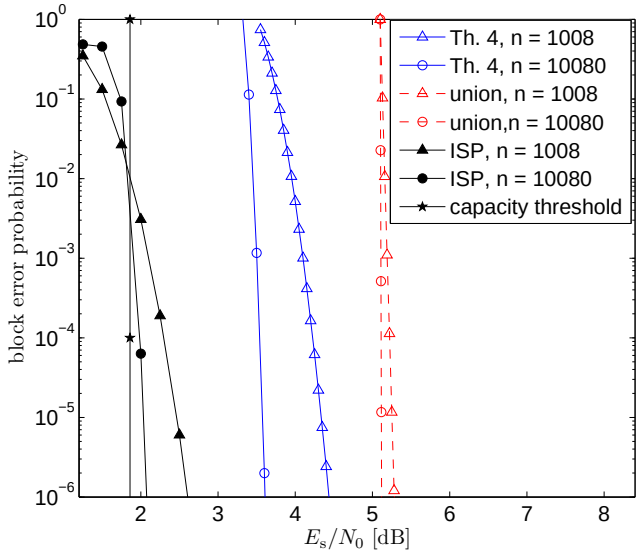


Figure 2.7: A comparison between the DS2 and union upper bounds on the block error probability under ML decoding for the $(8, 16)$-regular LDPC ensemble of Gallager (see Example 2.7). The transmission takes place over fully-interleaved Rayleigh fading channel with a QPSK modulation and perfect CSI at the receiver. The ISP lower bounds on the decoding error probability are shown for block lengths of 1008 and 10080 symbols. The capacity limit for infinite block length is also presented as a reference.

2.4 Summary and Conclusions

This chapter considers the performance of non-binary linear block codes whose transmission takes place over memoryless symmetric channels. To this end, upper bounds on the decoding error probability are derived for finite-length codes. The general bounding approach is based on a partitioning of the original ensemble into two subsets of codebooks, according to their minimal Hamming distance: The performance of the set of codebooks with a relatively low minimum Hamming distance is assessed via a simple union bound which only depends on the considered ensemble, whereas the other set is evaluated using the second version of the Duman and Salehi (DS2) bound (See Section 2.2.1). As a particular case of this bounding technique, an adaptation of the Shulman-Feder bound (SFB) (see [100]) is provided for non-binary linear block codes. The latter approach which is related to the adaptation of the SFB to the non-binary setting is similar to the work of Bennatan and Burshtein [11] for a different setting of coding with a random coset mechanism. Under a symmetry property of the ensemble, the resulting bound is considerably simplified and even tightened. This simplifying assumption, which holds in particular for the considered non-binary low-density parity-check (LDPC) ensembles, yields a bound whose summations are over

the Hamming weights of the non-zero codewords rather than their compositions (see Theorem 2.3). The tightness of the bounds presented in this chapter is exemplified for the non-binary regular LDPC ensembles of Gallager [44] where transmission takes place over the q-ary symmetric channel and the AWGN channel with a q-ary PSK modulation. The bound provided in Theorem 2.3 is attractive and show meaningful results up to the ultimate capacity limit. In addition, it outperforms the adaptation of the SFB in Theorem 2.2 for the non-binary setting which is even pronounced as the cardinality of the code alphabet is increased.

The weakness of the union bound is exemplified in this chapter for regular LDPC code ensembles, showing the necessity in the replacement of the union bound with some improved upper bounds on the decoding error probability. On the other hand, the bound provided in Theorem 2.3 is most attractive and shows meaningful results at a significant portion of the rate region between the cutoff rate and the ultimate channel capacity. The upper bound in Theorem 2.3 is compared to two lower bounds on the ultimate error performance of finite-length block codes (which hold for general block codes, either linear or non-linear): The 1959 sphere-packing (SP59) lower bound of Shannon [99], and the lower bound derived in [115]. These comparisons show by examples that recent sphere-packing bounds form a useful analytical tool for finite-length block codes.

Appendices

2.A Proof of Lemma 2.1

Let $x_1, x_2, x_3 \in \mathcal{X}$, p be the transition probability of the channel, and $\mathcal{T}$ be the mapping as in Lemma 2.1. Then, by setting $x \triangleq x_3 - x_2$, it follows from (2.1) that for all $y' \in \mathcal{Y}$

$$p(y'|x) = p\big(\mathcal{T}(y', x_2)|x_2 + x\big).$$

As a particular case, for $y' = \mathcal{T}(y, x_1)$ where $y \in \mathcal{Y}$, we have

$$p\big(\mathcal{T}(y, x_1)|x\big) = p\Big(\mathcal{T}\big(\mathcal{T}(y, x_1), x_2\big)|x_2 + x\Big). \tag{2.A.1}$$

Using (2.1) (repeatedly twice) on the LHS of (2.A.1) it follows that

$$p\big(\mathcal{T}(y, x_1)|x\big) = p(y|x - x_1) = p\big(\mathcal{T}(y, x_3 - x + x_1)|x_3\big). \tag{2.A.2}$$

which then yields from (2.A.1) and (2.A.2), jointly with the equality $x_3 - x = x_2$, that

$$p(\mathcal{T}(y, x_1 + x_2)|x_3) = p\Big(\mathcal{T}\big(\mathcal{T}(y, x_1), x_2\big)|x_3\Big)$$

which coincides with (2.2).

2.B Proof of Proposition 2.1

The following proof holds for channels with a discrete-output alphabet, and the generalization of the proof to continuous-output alphabet channels is trivial. Let p be the symmetric transition probability function of the considered channel, and $\mathcal{T}$ be its corresponding function according to Definition 2.1. The conditional error probability of the m-th message, $\mathbf{x}_m = (x_{m,1}, x_{m,2}, \ldots, x_{m,n})$, under ML decoding is given by

$$\begin{aligned} P_{\mathrm{e}|m} &= \sum_{\mathbf{y}\in\Lambda_m^c} \prod_{i=1}^{n} p\left(y_i|x_{m,i}\right) = \sum_{\mathbf{y}\in\Lambda_m^c} \prod_{x\in\mathcal{X}} \prod_{\{i:\ x_{m,i}=x\}} p(y_i|x) \\ &= \sum_{\mathbf{y}\in\Lambda_m^c} \prod_{x\in\mathcal{X}} \prod_{\{i:\ x_{m,i}=x\}} p(\mathcal{T}\left(y_i,-x\right)|0) \end{aligned}$$

where $\mathbf{y} = (y_1, \ldots, y_n)$, and

$$\begin{aligned} \Lambda_m^c &= \left\{ \mathbf{y} : \sum_{i=1}^{n} \ln\left(\frac{p(y_i|x_{m',i})}{p(y_i|x_{m,i})}\right) \geq 0,\ \text{for some } m' \neq m \right\} \\ &= \left\{ \mathbf{y} : \sum_{\{x,x'\in\mathcal{X}:\ x'\neq x\}} \sum_{\left\{i:\ x_{m',i}=x',x_{m,i}=x\right\}} \ln\left(\frac{p(y_i|x')}{p(y_i|x)}\right) \geq 0,\ \text{for some } m' \neq m \right\} \\ &= \left\{ \mathbf{y} : \sum_{\{x,x'\in\mathcal{X}:\ x'\neq x\}} \sum_{\left\{i:\ x_{m',i}=x',x_{m,i}=x\right\}} \ln\left(\frac{p(\mathcal{T}(y_i,-x')|0)}{p(\mathcal{T}(y_i,-x)|0)}\right) \geq 0,\ \text{for some } m' \neq m \right\}. \end{aligned}$$

Using the change of variables

$$z_i = \mathcal{T}(y_i, -x_{m,i}), \quad 1 \leq i \leq n$$

it follows that

$$P_{\mathrm{e}|m} = \sum_{\mathbf{z}\in\tilde{\Lambda}_m^c} \prod_{i=1}^{n} p(z_i|0)$$

where

$$\begin{aligned} &\tilde{\Lambda}_m^c \\ &= \left\{ \mathbf{z} : \sum_{\{x,x'\in\mathcal{X}:\ x'\neq x\}} \sum_{\left\{i:\ x_{m',i}=x',x_{m,i}=x\right\}} \ln\left(\frac{p(\mathcal{T}(z_i,x-x')|0)}{p(z_i|0)}\right) \geq 0,\ \text{for some } m' \neq m \right\} \\ &= \left\{ \mathbf{z} : \sum_{\delta\in\mathcal{X}} \sum_{\left\{i:\ x_{m,i}-x_{m',i}=\delta\right\}} \ln\left(\frac{p(\mathcal{T}(z_i,\delta)|0)}{p(z_i|0)}\right) \geq 0,\ \text{for some } m' \neq m \right\}. \end{aligned}$$

Since the code $\mathcal{C}$ is a linear space, then for every two codewords $\mathbf{x}_{m'} \neq \mathbf{x}_m$ in $\mathcal{C}$, there exists a third non-zero codeword $\mathbf{x}_l$ in $\mathcal{C}$ where $\mathbf{x}_l = \mathbf{x}_{m'} - \mathbf{x}_m$. Hence, for every $m = 1, 2, \ldots, M$ and for every $\mathbf{z} \in \tilde{\Lambda}_m^c$, there exists some $l \in \{1, 2, \ldots, M\}$ for which

$$\sum_{\delta \in \mathcal{X}} \sum_{\{i:\ -x_{l,i} = \delta\}} \ln\left(\frac{p(\mathcal{T}(z_i, \delta)|0)}{p(z_i|0)}\right) \geq 0.$$

Denote by $\mathbf{x}_1 \in \mathcal{C}$ the all-zero codeword, then it follows that

$$\tilde{\Lambda}_m^c = \tilde{\Lambda}_1^c, \quad m = 1, 2, \ldots, q^k$$

which concludes the proof.

2.C Proof of Proposition 2.2

Since the channel is symmetric, we have from Proposition 2.1 and (2.3) that

$$P_{\mathrm{e}} = P_{\mathrm{e}|0} \leq \left(\sum_{\mathbf{y} \in \mathcal{Y}^n} G_n^0(\mathbf{y}) p_n(\mathbf{y}|\mathbf{0})\right)^{1-\rho} \cdot \left\{\sum_{m' \neq 0} \sum_{\mathbf{y} \in \mathcal{Y}^n} G_n^0(\mathbf{y})^{1-\frac{1}{\rho}} p_n(\mathbf{y}|\mathbf{0}) \left(\frac{p_n(\mathbf{y}|\mathbf{x}_{m'})}{p_n(\mathbf{y}|\mathbf{0})}\right)^{\lambda}\right\}^{\rho}.$$

Next, setting $G_n^0(\mathbf{y})$ as in (2.4), for memoryless channels we have

$$P_{\mathrm{e}} \leq \left(\sum_{\mathbf{y} \in \mathcal{Y}^n} \prod_{i=1}^{n} g(y_i) p(y_i|0)\right)^{1-\rho} \cdot \left\{\sum_{m' \neq 0} \sum_{\mathbf{y} \in \mathcal{Y}^n} \prod_{i=1}^{n} g(y_i)^{1-\frac{1}{\rho}} p(y_i|0) \left(\frac{p(y_i|x_{m',i})}{p(y_i|0)}\right)^{\lambda}\right\}^{\rho}$$

which concludes the proof by replacing the sum of products with the corresponding product of sums.

2.D Proof of Theorem 2.2

From (2.8)

$$\Pr(\text{ error } \mid d_{\min} > D_n) \leq \left(\sum_{y \in \mathcal{Y}} g(y) p(y|0)\right)^{n(1-\rho)} q^{-n\rho(1-R)} \cdot \left(\sum_{\mathbf{t} \in \mathcal{H}:\ n - t_0 > D_n} \frac{\mathsf{E}\left[|\mathcal{C}_{\mathbf{t}}| \;\middle|\; d_{\min} > D_n\right]}{q^{-n(1-R)} \binom{n}{\mathbf{t}}} \binom{n}{\mathbf{t}} \prod_{x \in \mathcal{X}} \left(s_{\lambda,\rho}(x)\right)^{t_x}\right)^{\rho}$$

$$\leq \left(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\right)^{n(1-\rho)} q^{-n\rho(1-R)} \cdot \left(\max_{\mathbf{t}\in\mathcal{H}:\ n-t_0>D_n} \left\{ \frac{\mathsf{E}\left[|\mathcal{C}_{\mathbf{t}}| \;\middle|\; d_{\min} > D_n\right]}{q^{-n(1-R)}\binom{n}{\mathbf{t}}} \right\}\right)^{\rho}$$
$$\cdot \left(\sum_{\mathbf{t}\in\mathcal{H}:\ n-t_0>D_n} \binom{n}{\mathbf{t}} \prod_{x\in\mathcal{X}} \big(s_{\lambda,\rho}(x)\big)^{t_x}\right)^{\rho}$$

where the last transition holds since $\sum_i x_i y_i \leq \max_i x_i \sum_i y_i$ if $\{x_i\}$ and $\{y_i\}$ are non-negative sequences. Let $\mathcal{X}^* \triangleq \mathcal{X}\setminus\{0\}$, from the definition of α_q in (2.13) we get

$$\begin{aligned}
&\Pr(\text{ error } \mid d_{\min} > D_n) \\
&\quad\leq q^{-n\rho(1-R)} \left(\alpha_q(\mathcal{C}, D_n)\right)^{\rho} \left(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\right)^{n(1-\rho)} \\
&\qquad\cdot \left[\sum_{l=D_n+1}^{n} \binom{n}{l} \big(s_{\lambda,\rho}(0)\big)^{n-l} \sum_{t_1+\ldots+t_{q-1}=l} \binom{l}{t_1,\ldots,t_{q-1}} \prod_{x\in\mathcal{X}^*} \big(s_{\lambda,\rho}(x)\big)^{t_x}\right]^{\rho} \\
&\quad= q^{-n\rho(1-R)} \left(\alpha_q(\mathcal{C}, D_n)\right)^{\rho} \left(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\right)^{n(1-\rho)} \\
&\qquad\cdot \left[\sum_{l=D_n+1}^{n} \binom{n}{l} \big(s_{\lambda,\rho}(0)\big)^{n-l} \left(\sum_{x\in\mathcal{X}^*} s_{\lambda,\rho}(x)\right)^{l}\right]^{\rho}.
\end{aligned}$$

Consequently,

$$\begin{aligned}
&\Pr(\text{ error } \mid d_{\min} > D_n) \\
&\quad\leq q^{-n\rho(1-R)} \left(\alpha_q(\mathcal{C}, D_n)\right)^{\rho} \left(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\right)^{n(1-\rho)} \left(\sum_{x\in\mathcal{X}} s_{\lambda,\rho}(x)\right)^{n\rho}. \qquad (2.D.3)
\end{aligned}$$

Next, setting

$$g(y) = \left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{\rho} p(y|0)^{-\frac{\rho}{1+\rho}}, \quad \lambda = \frac{1}{1+\rho} \qquad (2.D.4)$$

it follows that

$$\sum_{y\in\mathcal{Y}} g(y)p(y|0) = \sum_{y\in\mathcal{Y}} \left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{\rho} p(y|0)^{\frac{1}{1+\rho}}. \qquad (2.D.5)$$

In addition, plugging (2.D.4) in (2.9), we get

$$s_{\lambda,\rho}(x) = \sum_{y\in\mathcal{Y}} \left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{\rho-1} p(y|0)^{\frac{1}{1+\rho}} p(y|x)^{\frac{1}{1+\rho}}$$

which then implies from (2.D.5) that

$$\begin{aligned}\sum_{x\in\mathcal{X}} s_{\lambda,\rho}(x) &= q\sum_{y\in\mathcal{Y}}\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{\rho} p(y|0)^{\frac{1}{1+\rho}} \\ &= q\sum_{y\in\mathcal{Y}} g(y)p(y|0). \end{aligned} \tag{2.D.6}$$

From (2.D.3) and (2.D.6), it follows that

$$\Pr(\text{ error } \mid d_{\min} > D_n) \le q^{n\rho R}\left(\alpha_q(\mathcal{C}, D_n)\right)^{\rho}\left(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\right)^{n}. \tag{2.D.7}$$

To complete the proof, we need the following lemma:

Lemma 2.D.1 Setting $g(y)$ as in (2.D.4), the following equality follows for all ξ:

$$\sum_{y\in\mathcal{Y}} g(y)^{\xi}p(y|0) = \sum_{y\in\mathcal{Y}}\left[\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{\xi\rho}\cdot\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{1-\frac{\xi\rho}{1+\rho}}\right)\right]. \tag{2.D.8}$$

Proof: Since the channel is symmetric, then there exists a function $\mathcal{T}$, as in Definition 2.1, satisfying (2.1) and (2.2). As a result, setting $g(y)$ as in (2.D.4) we have

$$\begin{aligned}&\sum_{y\in\mathcal{Y}} g(y)^{\xi}p(y|0) \\ &\quad= \sum_{y\in\mathcal{Y}}\left(\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{\rho} p(y|0)^{-\frac{\rho}{1+\rho}}\right)^{\xi} p(y|0) \\ &\quad= \sum_{y\in\mathcal{Y}} p(y|0)^{1-\frac{\xi\rho}{1+\rho}}\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{\xi\rho} \\ &\quad\overset{(a)}{=} \frac{1}{q}\sum_{x'\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(y|0)^{1-\frac{\xi\rho}{1+\rho}}\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{\xi\rho} \\ &\quad\overset{(b)}{=} \frac{1}{q}\sum_{x'\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(\mathcal{T}(y,x')|x')^{1-\frac{\xi\rho}{1+\rho}}\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y|x)^{\frac{1}{1+\rho}}\right)^{\xi\rho} \\ &\quad\overset{(c)}{=} \frac{1}{q}\sum_{x'\in\mathcal{X}}\sum_{y'\in\mathcal{Y}} p(y'|x')^{1-\frac{\xi\rho}{1+\rho}}\left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(\mathcal{T}(y',-x')|x)^{\frac{1}{1+\rho}}\right)^{\xi\rho}\end{aligned}$$

where in (a) an additional variable is added, (b) is based on (2.1), and (c) follows since

$$p(\mathcal{T}(\mathcal{T}(y,x'),-x')|x) = p(y|x) \tag{2.D.9}$$

for all $x, x' \in \mathcal{X}$ and $y \in \mathcal{Y}$. Next, using the closure of the (finite) input alphabet, it follows that

$$\begin{aligned}
&\sum_{y\in\mathcal{Y}} g(y)^{\xi} p(y|0) \\
&\quad \overset{\text{(a)}}{=} \frac{1}{q}\sum_{x'\in\mathcal{X}}\sum_{y'\in\mathcal{Y}} p(y'|x')^{1-\frac{\xi\rho}{1+\rho}} \left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(\mathcal{T}(\mathcal{T}(y',-x'),x+x'-x)|x+x')^{\frac{1}{1+\rho}}\right)^{\xi\rho} \\
&\quad \overset{\text{(b)}}{=} \frac{1}{q}\sum_{x'\in\mathcal{X}}\sum_{y'\in\mathcal{Y}} p(y'|x')^{1-\frac{\xi\rho}{1+\rho}} \left(\frac{1}{q}\sum_{x\in\mathcal{X}} p(y'|x+x')^{\frac{1}{1+\rho}}\right)^{\xi\rho} \\
&\quad = \frac{1}{q}\sum_{x'\in\mathcal{X}}\sum_{y'\in\mathcal{Y}} p(y'|x')^{1-\frac{\xi\rho}{1+\rho}} \left(\frac{1}{q}\sum_{x''\in\mathcal{X}} p(y'|x'')^{\frac{1}{1+\rho}}\right)^{\xi\rho} \\
&\quad = \sum_{y'\in\mathcal{Y}} \left[\left(\frac{1}{q}\sum_{x'\in\mathcal{X}} p(y'|x')^{1-\frac{\xi\rho}{1+\rho}}\right)\cdot\left(\frac{1}{q}\sum_{x''\in\mathcal{X}} p(y'|x'')^{\frac{1}{1+\rho}}\right)^{\xi\rho}\right]
\end{aligned}$$

where (a) follows from (2.1) and (b) follows from (2.2) and (2.D.9), both with $x_1 = x$ and $x_2 = x + x'$. This concludes the proof. ■

From (2.D.7) and Lemma 2.D.1 (with $\xi = 1$ in (2.D.8)), we get after an optimization over ρ (where $0 \le \rho \le 1$):

$$\Pr(\text{ error } \mid d_{\min} > D_n) \le q^{-nE_{\mathrm{r}}\left(R+\frac{\log_q \alpha_q(\mathcal{C},D_n)}{n}\right)}. \tag{2.D.10}$$

Finally, the proof of Theorem 2.2 follows from Lemma 2.3 and (2.D.10).

2.E Proof of Theorem 2.3

Under the conditions in Theorem 2.3, we get from (2.8), (2.14), and (2.15) that

$$\begin{aligned}
\Pr(\text{ error } \mid d_{\min} > D_n) &\le \left(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\right)^{n(1-\rho)} \\
&\cdot \left[\sum_{n-t_0>D_n} \frac{P(n-t_0)}{1-\epsilon_n}\binom{n}{t_0}\left(s_{\lambda,\rho}(0)\right)^{t_0} \sum_{t_1+\ldots+t_{q-1}=n-t_0}\binom{n-t_0}{t_1,\ldots,t_{q-1}}\prod_{x\in\mathcal{X}^*}\left(s_{\lambda,\rho}(x)\right)^{t_x}\right]^{\rho} \\
&= \left(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\right)^{n(1-\rho)} \cdot \left[\sum_{n-t_0>D_n} \frac{P(n-t_0)}{1-\epsilon_n}\binom{n}{t_0}\left(s_{\lambda,\rho}(0)\right)^{t_0}\left(\sum_{x\in\mathcal{X}^*} s_{\lambda,\rho}(x)\right)^{n-t_0}\right]^{\rho}
\end{aligned}$$

where $\mathcal{X}^* \triangleq \mathcal{X} \setminus \{0\}$. Next, setting λ and $g(y)$ as defined in (2.D.4), then it follows from (2.D.6) that

$$\Pr(\text{ error } \mid d_{\min} > D_n) \leq \left(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\right)^{n(1-\rho)} \cdot \left[\sum_{n-t_0>D_n} \frac{P(n-t_0)}{1-\epsilon_n}\binom{n}{t_0}\left(s_{\lambda,\rho}(0)\right)^{t_0}\left(q\sum_{y\in\mathcal{Y}} g(y)p(y|0) - s_{\lambda,\rho}(0)\right)^{n-t_0}\right]^{\rho}. \tag{2.E.11}$$

The proof is completed by applying Lemma 2.D.1 in (2.E.11) with $\xi = 1$ for

$$\sum_{y\in\mathcal{Y}} g(y)p(y|0)$$

and with $\xi = 1 - \frac{1}{\rho}$ for

$$s_{\lambda,\rho}(0) = \sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}}p(y|0).$$

2.F Proof of Lemma 2.5

Denote by $a_{x^*}(l)$ the number of choices of l not necessarily distinct non-zero elements in $\{1,\ldots,q-1\}$ whose summation modulo q equals x^* (where $x^* \in \{0,\ldots,q-1\}$). Then, for $1 \leq l \leq d$, there are $\binom{d}{l}a_{x^*}(l)$ vectors $\mathbf{x} = (x_1,\ldots,x_d)$, whose Hamming weight is l, and which satisfy

$$x_1 + \cdots + x_d = x^* \bmod q.$$

The sequences $\{a_{x^*}(l)\}$ satisfy the following system of recursive equations:

$$a_{x^*}(l) = \sum_{x=1}^{q-1} a_{(x^*-x) \bmod q}(l-1), \quad x^* = 0,1,\ldots,q-1 \tag{2.F.12}$$

with the initial conditions $a_0(1) = 0$, and $a_x(1) = 1$ for $x \in \{1,\ldots,q-1\}$. Using a vector notation, the equations in (2.F.12) are written as

$$\begin{pmatrix} a_0(l) \\ a_1(l) \\ \vdots \\ \\ a_{q-1}(l) \end{pmatrix} = \begin{pmatrix} 0 & 1 & \cdots & 1 & 1 \\ 1 & 0 & 1 & \cdots & 1 \\ \vdots & & \ddots & & \\ 1 & \cdots & 1 & 0 & 1 \\ 1 & & \cdots & 1 & 0 \end{pmatrix}_{q\times q} \begin{pmatrix} a_0(l-1) \\ a_1(l-1) \\ \vdots \\ \\ a_{q-1}(l-1) \end{pmatrix}$$

whose solution for $l \geq 1$ is given by

$$\begin{pmatrix} a_0(l) \\ a_1(l) \\ \vdots \\ \\ a_{q-1}(l) \end{pmatrix} = \begin{pmatrix} 0 & 1 & \cdots & 1 & 1 \\ 1 & 0 & 1 & \cdots & 1 \\ \vdots & & \ddots & & \\ 1 & \cdots & 1 & 0 & 1 \\ 1 & & \cdots & 1 & 0 \end{pmatrix}_{q\times q}^{l-1} \begin{pmatrix} 0 \\ 1 \\ \vdots \\ \\ 1 \end{pmatrix}_{q\times 1} . \tag{2.F.13}$$

In proving the considered lemma, the main ingredient is obtaining the number of vectors $\mathbf{x}$ satisfying the parity-check equation

$$x_1 + \cdots + x_d = 0 \bmod q. \tag{2.F.14}$$

Accordingly, only the sequence $\{a_0(l)\}$ is of interest. To obtain a closed form expression for this sequence, consider the following difference equation:

$$\begin{cases} u_l = (q-1)\big(u_{l-1} + (-1)^l\big) \\ u_1 = 0 \end{cases} . \tag{2.F.15}$$

It can be verified by induction that the elements on the diagonal of the $q \times q$ matrix on the RHS of (2.F.13), raised to the $(l-1)$-th power, are identical and equal to u_{l-1}, where the sequence $\{u_l\}$ is the solution of (2.F.15). Moreover, all other elements outside the diagonal, are equal to $u_{l-1} + (-1)^l$. As a result, it follows from (2.F.13) that

$$a_0(l) = (q-1)\left(u_{l-1} + (-1)^l\right),\ l \geq 1,\ a_0(1) = 0.$$

which implies from (2.F.15) that $a_0(l) = u_l$ for $l \geq 1$. Solving the difference equation in (2.F.15), gives

$$a_0(l) = \frac{(q-1)^l + (q-1)(-1)^l}{q}, \quad l \geq 1.$$

Hence, the enumerator for the number of vectors $\mathbf{x}$ satisfying the parity-check equation in (2.F.14), is given by $A^*(X)$ in (2.21). As a result, the enumerator of the first submatrix in the considered ensemble is given in (2.20) (this is similar to the idea provided in [105] for the binary case). Finally, (2.19) is established in [44] which concludes the proof of Lemma 2.5.

2.G Proof of Theorem 2.4

Using the DS2 bound for the case at hand, it follows that

$$
\begin{aligned}
&P(\text{ error } | d_{\min} > D_n) \\
&= \mathsf{E}\left[\iint_{(\mathbf{y},\mathbf{a}):p(\mathbf{y},\mathbf{a}|\mathbf{x})\geq p(\mathbf{y},\mathbf{a}|\mathbf{0}) \text{ for some } \mathbf{x}\neq\mathbf{0}\in\mathcal{C}} p(\mathbf{y},\mathbf{a}|\mathbf{0})\ d\mathbf{y}\, d\mathbf{a} \,\middle|\, d_{\min} > D_n \right] \\
&\leq \mathsf{E}\left[\iint_{\mathbf{y},\mathbf{a}} p(\mathbf{y},\mathbf{a}|\mathbf{0}) \sum_{j=1}^{J} \left(\sum_{\mathbf{t}\in\mathcal{H}_j} \sum_{\mathbf{x}\in\mathcal{C}_\mathbf{t}} \left(\frac{p(\mathbf{y},\mathbf{a}|\mathbf{x})}{p(\mathbf{y},\mathbf{a}|\mathbf{0})} \right)^{\lambda_j} \right)^{\rho_j} d\mathbf{y}\, d\mathbf{a} \,\middle|\, d_{\min} > D_n \right] \\
&= \sum_{j=1}^{J} \mathsf{E}\Bigg[\iint_{\mathbf{y},\mathbf{a}} \psi_j(\mathbf{y},\mathbf{a}) \\
&\qquad \cdot \left(\sum_{\mathbf{t}\in\mathcal{H}_j} \sum_{\mathbf{x}\in\mathcal{C}_\mathbf{t}} \psi_j(\mathbf{y},\mathbf{a})^{-\frac{1}{\rho_j}} p(\mathbf{y},\mathbf{a}|\mathbf{0})^{\frac{1-\lambda_j\rho_j}{\rho_j}} p(\mathbf{y},\mathbf{a}|\mathbf{x})^{\lambda_j} \right)^{\rho_j} d\mathbf{y}\, d\mathbf{a} \,\Bigg|\, d_{\min} > D_n \Bigg]
\end{aligned}
\tag{2.G.16}
$$

where the statistical expectation is taken over all the codebooks whose Hamming minimum distance is larger than D_n. From (2.G.16), using Jensen's inequality we have

$$
\begin{aligned}
&P(\text{ error } | d_{\min} > D_n) \\
&\leq \sum_{j=1}^{J} \mathsf{E}\left[\left(\sum_{\mathbf{t}\in\mathcal{H}_j} \sum_{\mathbf{x}\in\mathcal{C}_\mathbf{t}} \iint_{\mathbf{y},\mathbf{a}} \psi_j(\mathbf{y},\mathbf{a})^{1-\frac{1}{\rho_j}} p(\mathbf{y},\mathbf{a}|\mathbf{0})^{\frac{1-\lambda_j\rho_j}{\rho_j}} p(\mathbf{y},\mathbf{a}|\mathbf{x})^{\lambda_j} \right)^{\rho_j} d\mathbf{y}\, d\mathbf{a} \,\middle|\, d_{\min} > D_n \right].
\end{aligned}
$$

Setting $\psi_j(\mathbf{y},\mathbf{a}) = \prod_i \psi_j(y_i,a_i)$, since the channel is memoryless we have

$$
\begin{aligned}
&P(\text{ error } | d_{\min} > D_n) \\
&\leq \sum_{j=1}^{J} \mathsf{E}\Bigg[\Bigg(\sum_{\mathbf{t}\in\mathcal{H}_j} \sum_{\mathbf{x}\in\mathcal{C}_\mathbf{t}} \\
&\qquad \iint_{\mathbf{y},\mathbf{a}} \prod_{i=1}^{n} \psi_j(y_i,a_i)^{-\frac{1}{\rho_j}} p(y_i,a_i|0)^{\frac{1-\lambda_j\rho_j}{\rho_j}} p(y_i,a_i|x_i)^{\lambda_j}\, dy_i\, da_i \Bigg)^{\rho_j} \,\Bigg|\, d_{\min} > D_n \Bigg] \\
&= \sum_{j=1}^{J} \mathsf{E}\Bigg[\Bigg(\sum_{\mathbf{t}\in\mathcal{H}_j} |\mathcal{C}_\mathbf{t}| \\
&\qquad \prod_{x\in\mathcal{X}} \left(\iint_{y,a} \psi_j(y,a)^{1-\frac{1}{\rho_j}} p(y,a|0)^{\frac{1-\lambda_j\rho_j}{\rho_j}} p(y,a|x)^{\lambda_j}\, dy\, da \right)^{t_x} \Bigg)^{\rho_j} \,\Bigg|\, d_{\min} > D_n \Bigg].
\end{aligned}
$$

The proof is concluded by using Jensen's inequality (for the statistical expectation) and Lemma 2.3.

2.H A Closed-form expression for the integral in Theorem 2.4 when applied to Example 2.7

Similarly to [58], we will pursue a closed-form expression by examining an exponential tilting probability measure ψ as in (2.26). Note that the joint pdf $p(y,a|x)$ to receive the noisy observation $y \in \mathbb{C}$ with a fading sample $a \geq 0$, given that the transmitted symbol is $x \in \mathcal{X}$, is given according to

$$p(y,a|x) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}\left|y - a\mu(x)\right|^2\right) p(a),$$

where

$$p(a) = 2(1+K)a \exp\Big(-(1+K)a^2 - K\Big) I_0\left(2a\sqrt{K(K+1)}\right), \quad a \geq 0,$$

is the pdf of the Rician fading sample $a \in \mathcal{A}$ with a parameter K, and $\mu(x) \triangleq \sqrt{\frac{2E_\mathrm{s}}{N_0}} \exp\left(\frac{2\pi i}{q}\, x\right)$ is the q-ary PSK modulation mapping applied in the considered scheme. In addition, ψ_j in (2.26) is easily verified to be a probability measure. Assuming that $1 + K + \beta > 0$ (which is the case since $\alpha \geq 0$), the denominator of ψ as in (2.26) equals

$$\int_0^\infty p(a) \exp\left(-\frac{\alpha v^2 a^2 E_\mathrm{s}}{N_0}\right) da = \frac{1+K}{1+K+\beta} \exp\left(-\frac{\beta K}{1+K+\beta}\right).$$

Straightforward (though tedious) calculations show that for every $x \in \mathcal{X}$

$$\begin{aligned}
&\int_{a=0}^\infty \int_{y\in\mathbb{C}} \psi(y,a)^{1-\frac{1}{\rho}} p(y,a|0)^{\frac{1-\lambda\rho}{\rho}} p(y,a|x)^\lambda \, dy\, da \\
&\qquad = \frac{\rho}{1-\alpha(1-\rho)} \left(\frac{1+K}{\alpha(1+K+\beta)} \exp\left(-\frac{\beta K}{1+K+\beta}\right)\right)^{\frac{1}{\rho}-1} \\
&\qquad\quad \cdot \frac{1+K}{1+K+\gamma_x} \exp\left(-\frac{\gamma_x K}{1+K+\gamma_x}\right)
\end{aligned}$$

where

$$\begin{aligned}
\beta &\triangleq \frac{\alpha v^2 E_\mathrm{s}}{N_0} \\
\gamma_x &\triangleq \beta\left(1-\frac{1}{\rho}\right) - \frac{\rho E_\mathrm{s}}{\big(1-\alpha(1-\rho)\big)N_0} \left|\alpha u\left(1-\frac{1}{\rho}\right) + \frac{1-\lambda\rho}{\rho} + \lambda e^{\frac{2\pi i}{q}x}\right|^2 \\
&\qquad + \frac{E_\mathrm{s}}{N_0}\left(\alpha\left|u\right|^2\left(1-\frac{1}{\rho}\right) + \frac{1}{\rho}\right).
\end{aligned}$$

2.I A Closed-form expression for the integral in Theorem 2.4 when applied to Example 2.8

The following exponential tilting measure is applied:

$$\psi(y,a) = \frac{\alpha p(a)}{2\pi}\left(1+\frac{\beta}{L}\sqrt{\frac{2E_s}{N_0}}\right)^L \exp\left(-\frac{\alpha}{2}\left|y - a\sqrt{\frac{2E_{\mathrm{s}}}{N_0}}u\right|^2 - \beta a^2\sqrt{\frac{2E_{\mathrm{s}}}{N_0}}\right) \quad (2.\mathrm{I}.17)$$

where y is complex-valued, $a, \alpha, \beta \geq 0$, are real-valued parameters, u is a complex-valued parameter, and $p(a)$ is the pdf of the fading, given in (2.27). The integral in (2.24) with the proposed tilting measure in (2.I.17) is calculated via straightforward calculus, and it is obtained that for every $x \in \mathcal{X}$

$$\begin{aligned}
&\int_{a=0}^{\infty}\int_{y\in\mathbb{C}} \psi(y,a)^{1-\frac{1}{\rho}} p(y,a|0)^{\frac{1-\lambda\rho}{\rho}} p(y,a|x)^{\lambda}\, dy\, da \\
&= \frac{\rho\alpha^{1-\frac{1}{\rho}}L^L}{1-\alpha\,(1-\rho)}\left(1+\frac{\beta}{L}\sqrt{\frac{2E_{\mathrm{s}}}{N_0}}\right)^{L\left(1-\frac{1}{\rho}\right)} \\
&\left(L+\beta\left(1-\frac{1}{\rho}\right)\sqrt{\frac{2E_{\mathrm{s}}}{N_0}} + \left(1-\frac{1}{\rho}\right)\frac{\alpha\left|u\right|^2 E_{\mathrm{s}}}{N_0} + \frac{E_{\mathrm{s}}}{\rho N_0}\right. \\
&\quad \left. - \frac{\rho E_{\mathrm{s}}}{N_0\,(1-\alpha\,(1-\rho))}\left|\alpha u\left(1-\frac{1}{\rho}\right) + \frac{1-\lambda\rho}{\rho} + \lambda\exp\left(\frac{2\pi i x}{q}\right)\right|^2\right)^{-L}.
\end{aligned}$$

Chapter 3

Performance Bounds for Erasure, List and Decision Feedback Schemes with Linear Block Codes

Chapter Overview

A message independence property and some new performance upper bounds are derived in this chapter for erasure, list and decision-feedback schemes with linear block codes transmitted over memoryless symmetric channels. Similar to the classical work of Forney, this chapter is focused on the derivation of some Gallager-type bounds on the achievable tradeoffs for these coding schemes, where the main novelty is the suitability of the bounds for both random and structured linear block codes (or code ensembles). The bounds are applicable to finite-length codes and to the asymptotic case of infinite block length, and they are applied to low-density parity-check code ensembles. The chapter is based on the following paper:

E. Hof, I. Sason, and S. Shamai (Shitz), "Performance Bounds for Erasure, List and Decision Feedback Schemes with Linear Block Codes," *IEEE Trans. on Information Theory*, vol. 56, no. 8, pp. 3754–3778, August 2010.

This chapter is structured as follows: The definitions generalized decoding rules, and some of their basic properties, are provided in Section 3.1. New upper bounds under the generalized decoding rules in [41] are derived in Section 3.2. Error performance of suboptimal decoding rules are provided in Sections 3.4 and 3.5. Section 3.6 concludes the discussion. Some technical details are relegated to the appendices.

3.1 Channel Symmetry, Generalized Decoding, and Message Independence

In this section we introduce some definitions, examples, and statements related to channel symmetry, Forney's generalized decoding rule [41], and sub-optimal versions ([9] and [41]), as well as list decoding rules ([36] and [117]). A message independence property is stated for these decoding rules, which is used for the simplification of the analysis. The notation in Section 2.1 is assumed. In addition, a memoryless symmetric channel is assumed (see Definition 2.1), whose transition probability (or probability density, respectively) function is denoted by $p(y|x)$, where $x \in \mathcal{X}$ and $y \in \mathcal{Y}$.

Let $\mathcal{C} = \{\mathbf{x}_m\}_{m=1}^{q^k}$ be a linear block code whose generator matrix is a $k \times n$ full-rank matrix with entries over $\mathcal{X}$. The decoding rules studied in this chapter are specified in terms of decision regions Λ_m, $1 \leq m \leq q^k$, which are all subsets of $\mathcal{Y}^n$. The conditional error probability of the m-th message is given by

$$P_{\mathrm{e}|m} = \sum_{\mathbf{y} \in \Lambda_m^{\mathrm{c}}} p(\mathbf{y}|\mathbf{x}_m) \tag{3.1}$$

where Λ_m forms the decision region for the m-th codeword, and the superscript 'c' stands for the complementary set. The decision region of the m-th codeword under ML decoding gets the form

$$\Lambda_m = \left\{\mathbf{y} : p(\mathbf{y}|\mathbf{x}_m) > p(\mathbf{y}|\mathbf{x}_{m'}),\ \forall\, m' \neq m\right\} \tag{3.2}$$

where ties are resolved randomly with equal probability. Assuming equal a-priori probabilities for the transmitted messages, the ML decoding rule minimizes the error probability given in (3.1). A well-known result for binary linear block codes operating over MBIOS channels is that their error probability under ML decoding is independent of the transmitted codeword. This enables a great simplification in the analysis by assuming that the all-zero codeword is transmitted. This result is generalized in Chapter 2 for non-binary linear block codes whose transmission takes place over memoryless symmetric channels with discrete input alphabet.

When generalized decoding rules are considered, the decision regions Λ_m are not necessarily disjoint nor they include all the possible received vectors. The former case corresponds to decoding rules with a possibly *variable* list-size, and the latter case corresponds to decoding with erasures. A list is produced by the decoder where the received vector may possibly belong to more than one decision region. An erasure

event is declared by the decoder when the received vector does not belong to any decision region. These concepts were first introduced in [41]. When generalized decoding rules are allowed, the conditional block error probability $P_{\mathrm{e}|m}$ in (3.1) stands for the probability of either an undetected error or an erasure. When the decision regions are disjoint, the conditional undetected error probability is given by

$$P_{\mathrm{ue}|m} = \sum_{m' \neq m} \sum_{\mathbf{y} \in \Lambda_{m'}} p(\mathbf{y}|\mathbf{x}_m). \tag{3.3}$$

In addition, let $P_{\mathrm{x}|m}$ denote the conditional probability of an erasure event given that $\mathbf{x}_m$ is transmitted. Then

$$P_{\mathrm{x}|m} = P_{\mathrm{e}|m} - P_{\mathrm{ue}|m}.$$

In the case where list decoding is considered, the decision regions are not disjoint, and $P_{\mathrm{ue}|m}$ as given in (3.3) is no longer a probability. However the RHS of (3.3) equals the conditional expectation of the number of incorrect codewords in the list (the same notation, $P_{\mathrm{ue}|m}$, is used in both cases to simplify the statement of the following results). The optimum decoding rule with respect to the tradeoff between the error and the undetected error event is derived in [41].

Definition 3.1 (Forney's generalized decoding) Consider a block code over an alphabet $\mathcal{X}$, and let $\{\mathbf{x}_m\}$ denote its codebook. The generalized decoding rule is defined by the following decision regions:

$$\Lambda_m = \left\{ \mathbf{y} \in \mathcal{Y}^n : \ \frac{\Pr(\mathbf{y}, \mathbf{x}_m)}{\sum_{m' \neq m} \Pr(\mathbf{y}, \mathbf{x}_{m'})} \geq e^{nT} \right\} \tag{3.4}$$

where m is the index of the codeword, $T \in \mathbb{R}$ is a parameter, $\Pr(\mathbf{y}, \mathbf{x}_m)$ denotes the joint probability that $\mathbf{x}_m$ is the transmitted codeword and $\mathbf{y}$ is the received vector, and the summation is over all codewords except for $\mathbf{x}_m$.

Remark 3.1 The decision region in (3.4) can be expressed equivalently in the form

$$\Lambda_m = \left\{ \mathbf{y} \in \mathcal{Y}^n : \ \Pr(\mathbf{x}_m|\mathbf{y}) \geq \frac{e^{nT}}{1 + e^{nT}} \right\}. \tag{3.5}$$

Note that for $T = 0$, this decision region includes all the vectors $\mathbf{y} \in \mathcal{Y}^n$ for which $\Pr(\mathbf{x}_m|\mathbf{y}) \geq \frac{1}{2}$. The a-posteriori probability of $\mathbf{x}_m$, given that $\mathbf{y} \in \Lambda_m$ is received, is therefore larger than the a-posteriori probability for any other codeword. Hence, if a codeword is selected according to the decoder with the decision regions in (3.5) with $T = 0$, then the same decision is made by a MAP decoder (as no other codeword can get an a-posteriori probability larger than $\frac{1}{2}$). This implies that the undetected

error exponent for the decoder in (3.5) with $T = 0$ cannot be smaller than the error exponent of an ML decoder with equally-likely codewords. Interestingly, as will be shown later, we get the same lower bound on the error exponents for both decoders. Moreover, it is shown that for $T = 0$ the bounds for the undetected error event and erasures coincides.

Remark 3.2 The threshold parameter T in (3.4) controls the tradeoff between erasures and undetected errors (or average list size and decoding error). Setting $T > 0$ guarantees that the decision regions Λ_m are disjoint.

Proposition 3.1 (Forney's generalized decoding [41]) Assume that the decoding of a block code is carried according to the generalized decoding rule in Definition 3.1. Then, there is no other decoding rule that simultaneously gives a lower error probability and a lower undetected error probability (or an average number of incorrect codewords when list decoding is considered).

The following proposition generalizes the message independence property for the case of generalized decoding:

Proposition 3.2 (Message independence property for optimal generalized decoding) Let $\mathcal{C}$ be a linear block code whose transmission takes place over a memoryless and symmetric channel. Then, the block error probability and the undetected error probability, under the generalized decoding rule in Definition 3.1, are independent of the transmitted codeword.

Proof: See Appendix 3.A. ■

Remark 3.3 In the case where list decoding is considered (i.e., the decision regions are not disjoint), then Proposition 3.2 holds when we refer to the conditional expectation of the number of incorrect messages in the list produced by the generalized decoding rule, instead of the undetected error probability.

The following suboptimal decoding rule is suggested in [41] for the case of decoding with erasures:

Definition 3.2 (Likelihood Ratio (LR) Decoding) Consider a block code over the alphabet $\mathcal{X}$, and let $\{\mathbf{x}_m\}$ denote its codebook. The LR decoding rule is defined by the following decision regions:

$$\Lambda_m^{\mathrm{LR}} = \left\{ \mathbf{y} \in \mathcal{Y}^n : \ \frac{\Pr(\mathbf{y}, \mathbf{x}_m)}{\Pr(\mathbf{y}, \mathbf{x}_{m_2})} \geq e^{nT} \right\} \tag{3.6}$$

where m is a codeword index, $T > 0$ is a parameter, $\Pr(\mathbf{y}, \mathbf{x}_m)$ denotes the joint probability that $\mathbf{x}_m$ is the transmitted codeword and $\mathbf{y}$ is the received vector, and $m_2 = m_2(\mathbf{y})$ denotes the second most probable codeword for each received vector $\mathbf{y}$.

Remark 3.4 It is observed in [41] that the LR decoding rule may be a good approximation to the optimal regions in (3.4), since the second most likely codeword is usually much more probable than the rest of the codewords (excluding the most probable codeword). It is also noted in [41] that this suboptimal decoding rule is of practical utility.

Example 3.1 (Suboptimal generalized decoding) Consider the transmission of a binary linear block code over a BSC. Given a received vector $\mathbf{y} \in \{0,1\}^n$, the decoded codeword is $\mathbf{x}$ if and only if

$$d_{\mathrm{H}}(\mathbf{x}', \mathbf{y}) - d_{\mathrm{H}}(\mathbf{x}, \mathbf{y}) > 2\tau n \tag{3.7}$$

for all codewords $\mathbf{x}' \neq \mathbf{x}$, where $d_{\mathrm{H}}(\mathbf{x}, \mathbf{y})$ denoted the Hamming distance between $\mathbf{x}$, and $\mathbf{y}$, and $\tau \geq 0$ is an arbitrary parameter. Otherwise, an erasure is declared. It is easily verified that this rule is a particular case of (3.6). The error exponents for this setting are studied in [9].

The following proposition obtains a message independence property for the suboptimal decoding rule in Definition 3.2:

Proposition 3.3 (Message independence property for (suboptimal) LR decoding) Let $\mathcal{C}$ be a linear block code whose transmission takes place over a memoryless and symmetric channel. Then, the block error probability and the undetected error probability, under the suboptimal decoding rule in (3.6), are independent of the transmitted codeword.

Proof: See Appendix 3.B. ■

The following definition considers list decoding with a fixed size. Such a decoding rule is based on a fixed size of the list (instead of a variable list size which characterizes the decoding rule in Definition 3.1 with $T < 0$).

Definition 3.3 (Fixed-size list-decoding) Consider a block code over an alphabet $\mathcal{X}$, and let $\{\mathbf{x}_m\}$ denote its codebook. Given a fixed list size L, the list-decoder is a mapping from the set of all possible received vectors $\mathcal{Y}^n$ to the set of all possible lists of L codewords. This mapping produces the list whose likelihoods are the highest

among all other codewords. That is, given a received vector $\mathbf{y}$, a codeword $\mathbf{x}_m$ is in the list if $p(\mathbf{y}|\mathbf{x}_m) > p(\mathbf{y}|\mathbf{x}_{m'})$ for all $m' \neq m$ except for at most $L-1$ other possible codewords.

Assuming that the codeword $\mathbf{x}_m$ is transmitted, a block error event is occurred by the fixed-size list-decoding rule in Definition 3.3, if the list produced by the decoder does not include the transmitted codeword $\mathbf{x}_m$. The following proposition is analogous to the message independence property in Propositions 3.2 and 3.3:

Proposition 3.4 (Message independence property for fixed-size list decoding) Let $\mathcal{C}$ be a linear block code whose transmission takes place over a memoryless and symmetric channel. Then, the block error probability, under the fixed-size list-decoding is independent of the transmitted codeword.

Proof: See Appendix 3.C. ■

3.2 Upper Bounds under optimal generalized decoding

The transmission of block codes (not necessarily linear) is first considered. In addition, throughout the chapter, all codewords are assumed to have a uniform a-priori probability.

Proposition 3.5 Consider the transmission of a code $\mathcal{C}$ with a block length n and M codewords, and let $p(\mathbf{y}|\mathbf{x})$ designate the transition probability of the channel where $\mathbf{x} \in \mathcal{C}$ is the transmitted codeword and $\mathbf{y} \in \mathcal{Y}^n$ is the received vector. Then, the conditional block error probability ($P_{\mathrm{e}|m}$) and the average undetected error probability (P_{ue}) under the generalized decoding rule in (3.4) satisfy

$$P_{\mathrm{e}|m} \leq e^{nsT} D_{\mathrm{B}}(m, G_n^m, s, \rho) \tag{3.8}$$

$$P_{\mathrm{ue}} \leq e^{n(s-1)T} \frac{1}{M} \sum_{m=1}^{M} D_{\mathrm{B}}(m, G_n^m, s, \rho) \tag{3.9}$$

where $0 \leq s \leq \rho \leq 1$ are real-valued parameters, G_n^m is an arbitrary non-negative function over $\mathcal{Y}^n$ which possibly depends on the codeword $\mathbf{x}_m$, $1 \leq m \leq M$, and

$$D_{\mathrm{B}}(m, G_n^m, s, \rho) \triangleq \left(\sum_{\mathbf{y}} G_n^m(\mathbf{y}) p(\mathbf{y}|\mathbf{x}_m) \right)^{1-\rho} \left(\sum_{m' \neq m} \sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}_m) G_n^m(\mathbf{y})^{1-\frac{1}{\rho}} \left(\frac{p(\mathbf{y}|\mathbf{x}_{m'})}{p(\mathbf{y}|\mathbf{x}_m)} \right)^{\frac{s}{\rho}} \right)^{\rho}. \tag{3.10}$$

Proof: See Appendix 3.D. ■

Remark 3.5 Bounds (3.8) and (3.9) in Proposition 3.5 may be considered as a generalization of the DS2 bound ([26], [96], [94]). In fact, setting $T = 0$ in (3.8) reproduces the DS2 bound under ML decoding. Note however that for $T = 0$, the decision regions in (3.4) do not coincide with those under ML decoding (e.g., in the former case there are erasures).

The following corollary is a particularization of Proposition 3.5 for fully random block codes whose transmission takes place over memoryless channels. The corollary reproduces the exponential upper bounds as in [41, Th. 2].

Corollary 3.1 (Random coding error exponents under optimum generalized decoding) Consider the transmission of block codes over a memoryless communication channel with a transition probability law p. Then, under the notation in Proposition 3.5, there exists a block code which simultaneously satisfies

$$P_{\mathrm{e}} \leq e^{-nE_1(R,T)} \tag{3.11}$$

$$P_{\mathrm{ue}} \leq e^{-nE_2(R,T)} \tag{3.12}$$

where $R = \ln M/n$ is the code rate (in nats per channel use),

$$E_1(R,T) \triangleq \max_{0\leq s\leq\rho\leq 1,\ q_X} \Big(E_0(s,\rho,q_X) - \rho R - sT \Big) \tag{3.13}$$

$$E_2(R,T) \triangleq E_1(R,T) + T$$

$$E_0(s,\rho,q_X) \triangleq -\ln \sum_{y\in\mathcal{Y}} \left\{ \left(\sum_{x\in\mathcal{X}} q_X(x) p(y|x)^{1-s} \right) \left(\sum_{x\in\mathcal{X}} q_X(x) p(y|x)^{\frac{s}{\rho}} \right)^{\rho} \right\} \tag{3.14}$$

and q_X is a probability distribution over $\mathcal{X}$.

Proof: See Appendix 3.E. ■

The bounds in Corollary 3.1 are derived in [41] without relying on tilting measures. The current derivation relies on the DS2 bound which makes use of tilting measures and Jensen's inequality. It is noted in [41] that setting $T = 0$ in Corollary 3.1, provides the random coding error exponent of Gallager [45]. Hence, as is mentioned in [41], the random coding error exponent is attainable not only under ML decoding, but also under the generalized decoding rule in (3.4) with $T = 0$. The following proposition is a particularization of Proposition 3.5 for linear block codes.

Proposition 3.6 Consider an (n, k) linear block code $\mathcal{C}$ whose transmission takes place over a memoryless symmetric channel. Assume that the channel input and output alphabets are $\mathcal{X}$ and $\mathcal{Y}$, respectively, and let p be the transition probability of the channel. Then, the block error probability P_{e} and the undetected error probability P_{ue} under the generalized decoding rule in (3.4), satisfy

$$P_{\mathrm{e}} \leq e^{nsT} D(g, s, \rho) \tag{3.15}$$

$$P_{\mathrm{ue}} \leq e^{-n(1-s)T} D(g, s, \rho) \tag{3.16}$$

where $g : \mathcal{Y} \to \mathbb{R}$ is an arbitrary non-negative real-valued function, $0 \leq s \leq \rho \leq 1$ are arbitrary parameters, and

$$D(g, s, \rho) \triangleq \left(\sum_{y \in \mathcal{Y}} g(y) p(y|0) \right)^{n(1-\rho)} \left(\sum_{m' \neq 0} \prod_{i=1}^{n} \sum_{y \in \mathcal{Y}} g(y)^{1-\frac{1}{\rho}} p(y|0) \left(\frac{p(y|x_{m',i})}{p(y|0)} \right)^{\frac{s}{\rho}} \right)^{\rho}. \tag{3.17}$$

Proof: See Appendix 3.F. ∎

Remark 3.6 When the decision regions are not disjoint (i.e., a list decoder is considered), P_{ue} in (3.16) does not denote a probability but the expected number of incorrect codewords in the decoded list. The block error probability P_e in (3.15) refers, in this case, to the list decoding error probability.

Remark 3.7 The parameters s and ρ in Proposition 3.6 may be chosen separately for the bounds in (3.15) and (3.16). However, the optimized choice of the two parameters is identical in both bounds (since they only differ in the multiplicative term e^{-nT}).

The mathematical structure of the bound provided in the following corollary is similar to the Shulman-Feder bound (SFB) in [100]. Because of this reason, this bound may be considered as a generalization of the SFB for the generalized decoding rule in (3.4). To simplify the notation, the corollary is provided for the case of a binary linear block code whose transmission takes place over an MBIOS channel (the generalization of the bounds to non-binary linear block codes is performed similarly to the approach in the proof of Theorem 2.2).

Corollary 3.2 Consider an (n, k) binary linear block code $\mathcal{C}$ whose transmission takes place over an MBIOS channel with a transition probability law p. Then, the block error probability P_{e} and the undetected error probability P_{ue} under the generalized decoding rule in (3.4) satisfy

$$P_{\mathrm{e}} \leq e^{-n\left(E(\rho,R,\mathcal{C}) - \frac{\rho T}{1+\rho}\right)} \tag{3.18}$$

$$P_{\mathrm{ue}} \leq e^{-n\left(E(\rho,R,\mathcal{C}) + \frac{T}{1+\rho}\right)} \tag{3.19}$$

where $0 \leq \rho \leq 1$ is an arbitrary real-valued parameter, $R \triangleq \left(\frac{k}{n}\right) \cdot \ln 2$ is the code rate (in nats per channel use),

$$E\left(\rho, R, \mathcal{C}\right) \triangleq E_0(\rho) - \rho\left(R + \frac{\ln(\alpha(\mathcal{C}))}{n}\right) \tag{3.20}$$

$$E_0\left(\rho\right) \triangleq -\ln\left(\sum_y \left(\frac{1}{2}p(y|0)^{\frac{1}{1+\rho}} + \frac{1}{2}p(y|1)^{\frac{1}{1+\rho}}\right)^{1+\rho}\right) \tag{3.21}$$

$$\alpha(\mathcal{C}) \triangleq \max_{1\leq i\leq n} \frac{|\mathcal{C}_i|}{2^{-(n-k)}\binom{n}{i}} \tag{3.22}$$

and $|\mathcal{C}_i|$ denotes the number of codewords whose Hamming weight is i.

Proof: Setting $s = \frac{\rho}{1+\rho}$, and

$$g(y) = \left(\frac{1}{2}p(y|0)^{\frac{1}{1+\rho}} + \frac{1}{2}p(y|1)^{\frac{1}{1+\rho}}\right)^{\rho} p(y|0)^{-\frac{\rho}{1+\rho}} \tag{3.23}$$

in the bounds of Proposition 3.6, the proof follows in the same way as in [94, Ch. 4.4.1]. ■

Remark 3.8 In the case where the performance of an ensemble of linear block codes is of interest, repeating the derivation of Corollary 3.2 leads to the same upper bounds as in (3.18) and (3.19), where the cardinality $|\mathcal{C}_i|$ in (3.22) is replaced with its statistical expectation over the considered ensemble, and the codebooks of this ensemble are chosen uniformly at random.

Example 3.2 (Error exponents of fully random binary linear block codes) Consider the transmission of fully random binary linear (n, k) block codes over a memoryless symmetric channel. For this particular case, the term $\alpha(\mathcal{C})$ in (3.22) equals 1. As a result, it follows from Corollary 3.2 that the exponent of the block error probability (including erasures and undetected errors), denoted by E_{e}, satisfies

$$E_{\mathrm{e}} \geq \max_{0\leq\rho\leq 1}\left(E_0(\rho) - \rho R - \frac{\rho T}{1+\rho}\right) \tag{3.24}$$

where $E_0(\rho)$ is defined in (3.21), R is the code rate (in nats per channel use), and T is the parameter of the generalized decoding rule in Definition 3.1. Setting $T = 0$ in (3.24) reproduces the (non-expurgated) random coding error exponent of Gallager [45]. This observation was first made by Forney for the ensemble of fully random block codes [41]. The undetected error exponent, denoted by E_{ue}, satisfies

$$E_{\mathrm{ue}} \geq T + \max_{0\leq\rho\leq 1}\left(E_0(\rho) - \rho R - \frac{\rho T}{1+\rho}\right).$$

The lower bounds on the two error exponents are shown in Figures 3.1 and 3.2 for the case of transmission over a BSC with a crossover probability of $p = 0.11$, and for a binary-input AWGN channel with $E_{\rm s}/N_0 = -2.8$ dB, respectively (both values refer to the capacity limit for a rate of one-half bits per channel use). The bounds are sketched as a function of the code rate (in nats per channel use). The lower bounds on the error exponents for the case of decoding with erasures ($T \geq 0$) are provided in Figures 3.1(a) and 3.2(a) for $T = 0, 0.025, 0.05, 0.1$ and 0.15. For the case of decoding with a variable list-size ($T < 0$), the lower bounds on the error exponents are provided in Figures 3.1(b) and 3.2(b) for $T = 0, -0.05$, and -0.1. In addition, lower bounds on the exponent $E_N \triangleq -(\ln N)/n$, where N is the number of incorrect codewords in the decoded list, are also provided for this case. Note that the exponent E_N is negative above some rate. The figures show the region for which the exponent E_N is non-negative; the negative part of E_N, for which an upper bound on the size of the decoded list grows exponentially with the block length, is removed.

Recall the definitions of vector compositions, and complete composition spectrum in Chapter 2 (see Definitions 2.3 and 2.4).

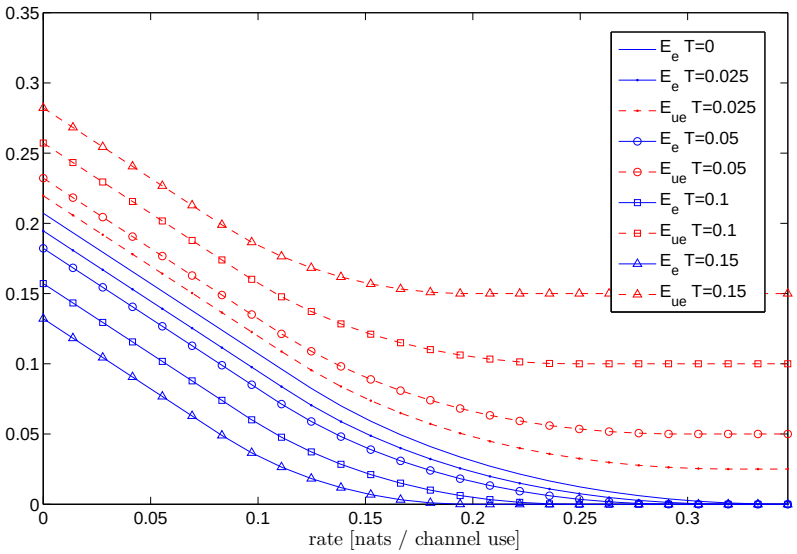
Corollary 3.3 Consider an ensemble $\mathcal{E}$ of (n, k) linear block codes whose transmission takes place over a memoryless symmetric channel. Let $P(l)$ denote the probability that a vector whose Hamming weight is l, forms a codeword in a randomly selected codebook from $\mathcal{E}$. Assume that the average composition spectrum over all the codes $\mathcal{C}$, uniformly selected at random from $\mathcal{E}$ satisfies

$$\mathsf{E}\Big[|\mathcal{C}_{\mathbf{t}}|\Big] = P(n - t_0)\binom{n}{\mathbf{t}}. \tag{3.25}$$

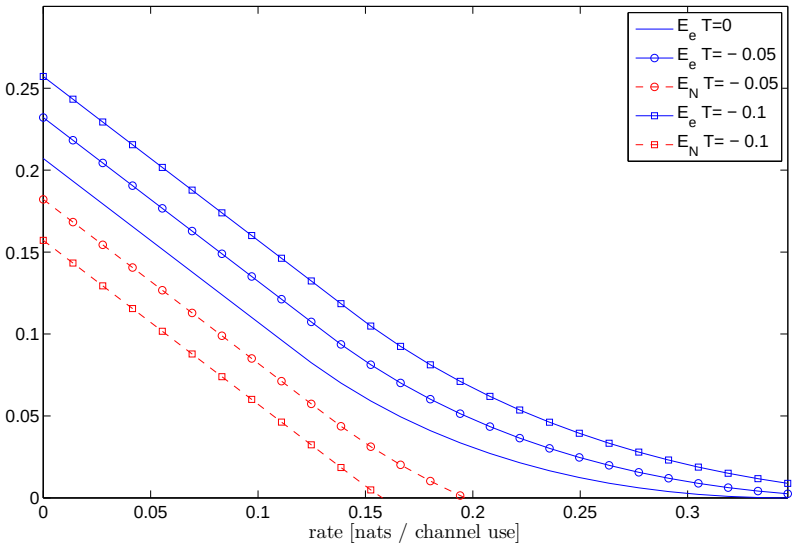
Then, under the notation in Proposition 3.6, the block error probability $P_{\rm e}$ and the undetected error probability $P_{\rm ue}$, satisfy

$$P_{\rm e} \leq e^{\frac{n\rho T}{1+\rho}} \cdot D_{\rm s}(\rho, \mathcal{C}) \tag{3.26}$$

$$P_{\rm ue} \leq e^{-\frac{nT}{1+\rho}} \cdot D_{\rm s}(\rho, \mathcal{C}) \tag{3.27}$$

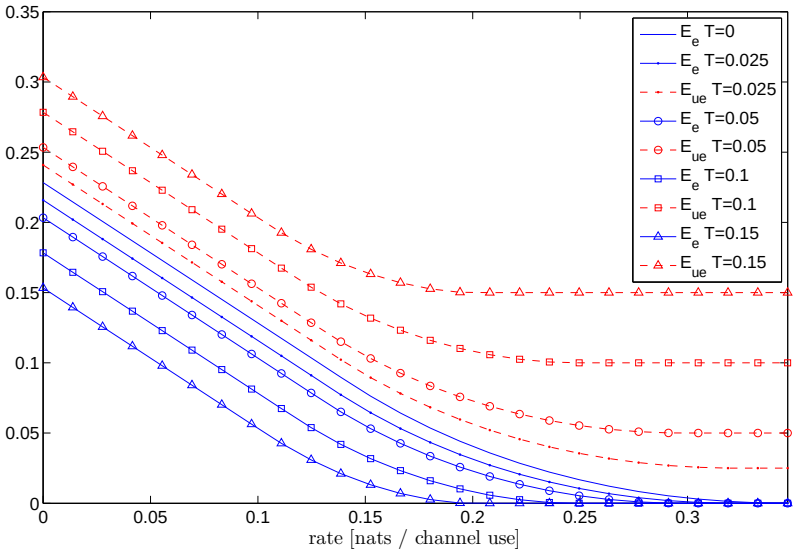


(a) Generalized decoding with erasures

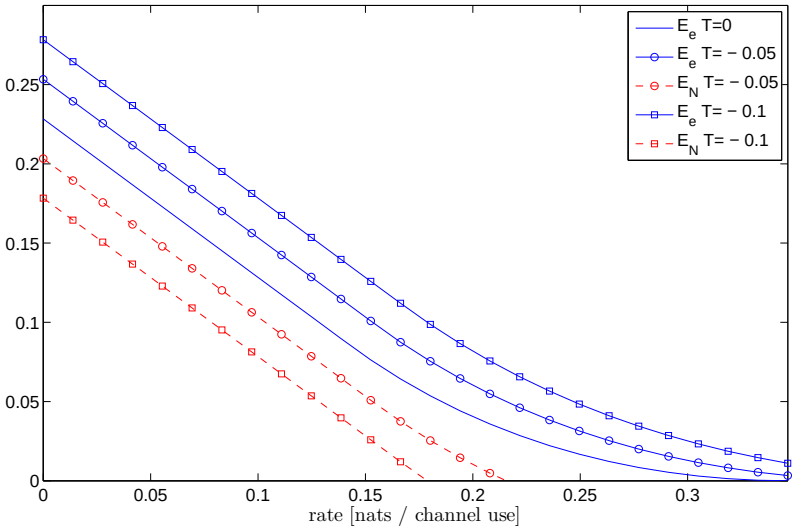


(b) Generalized decoding with a variable-size list

Figure 3.1: Lower bounds on the error exponents and list-size exponents for the ensemble of fully-random binary linear block codes whose transmission takes place over a BSC with a crossover probability of $p = 0.11$. The lower bounds in Corollary 3.2 are sketched in plots (a) and (b), for the generalized decoding rule in (3.4) with erasures (i.e., $T \geq 0$) and with a variable list-size (i.e., $T < 0$), respectively.



(a) Generalized decoding with erasures



(b) Generalized decoding with a variable-size list

Figure 3.2: Lower bounds on the error exponents and list-size exponents for the ensemble of fully-random binary linear block codes whose transmission takes place over a binary-input AWGN channel with $E_s/N_0 = -2.8$ dB. The lower bounds in Corollary 3.2 are sketched in plots (a) and (b), for the generalized decoding rule in (3.4) with erasures (i.e., $T \geq 0$) and with a variable list-size (i.e., $T < 0$), respectively.

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

$$D_{\mathrm{s}}(\rho, \mathcal{C}) \triangleq A(\rho)^{n(1-\rho)} \left(\sum_{1 \leq l \leq n} P(l) \binom{n}{l} B(\rho)^{n-l} C(\rho)^{l} \right)^{\rho} \tag{3.28}$$

$$A(\rho) \triangleq \sum_{y \in \mathcal{Y}} \left(\frac{1}{q} \sum_{x \in \mathcal{X}} p(y|x)^{\frac{1}{1+\rho}} \right)^{1+\rho} \tag{3.29}$$

$$B(\rho) \triangleq \sum_{y \in \mathcal{Y}} \left(\frac{1}{q} \sum_{x \in \mathcal{X}} p(y|x)^{\frac{1}{1+\rho}} \right)^{\rho-1} \left(\frac{1}{q} \sum_{x \in \mathcal{X}} p(y|x)^{\frac{2}{1+\rho}} \right) \tag{3.30}$$

$$C(\rho) \triangleq qA(\rho) - B(\rho). \tag{3.31}$$

Proof: Setting $s = \frac{\rho}{1+\rho}$ and choosing the tilting measure g in (3.23), the proof follows from Proposition 3.6 in the same way as in Theorem 2.3. ■

Remark 3.9 For an ensemble of *binary* linear block codes, the condition in (3.25) is not mandatory. Repeating the derivation results in the same bounds as in Corollary 3.3 where the term $P(l)\binom{n}{l}$ in (3.28) is replaced with the expected complete composition spectrum of the ensemble.

Remark 3.10 The bounds in Corollary 3.3 are tighter than those in Corollary 3.2. Hence, for a finite block length, the bounds in Corollary 3.3 are more attractive even though they lack the appealing exponential structure of the bounds in Corollary 3.2.

Remark 3.11 As a particular case of Remark 3.5, setting $T = 0$ in (3.26) reproduces the upper bound on the decoding error probability of non-binary linear block codes under ML decoding in Theorem 2.3.

The following comments concerns the numerical results shown in the examples throughout this chapter:

1. *Expurgation of codebooks*: The examples presented in this chapter consider the performance of some expurgated ensembles of regular LDPC codes under generalized decoding rules. Specifically, an expurgation of the codebooks whose minimum Hamming distance is not larger than a specific value D_n is assumed. As a result, the expected complete composition spectrum $\mathsf{E}\left[|\mathcal{C}_{\mathbf{t}}| \, | d_{\min} > D_n\right]$ of a codebook which is chosen uniformly at random from the expurgated ensemble, satisfies the following upper bound:

$$\mathsf{E}\big[|\mathcal{C}_{\mathbf{t}}| \, | d_{\min} > D_n\big] \leq \frac{\mathsf{E}\big[|\mathcal{C}_{\mathbf{t}}|\big]}{1 - \epsilon_n} \tag{3.32}$$

where $\mathsf{E}\big[|\mathcal{C}_{\mathbf{t}}|\big]$ is the expected composition spectrum of the original (non-expurgated) ensemble, and

$$\sum_{\mathbf{t}:\ n-t_0\leq D_n} \mathsf{E}\big[|\mathcal{C}_{\mathbf{t}}|\big] \leq \epsilon_n. \tag{3.33}$$

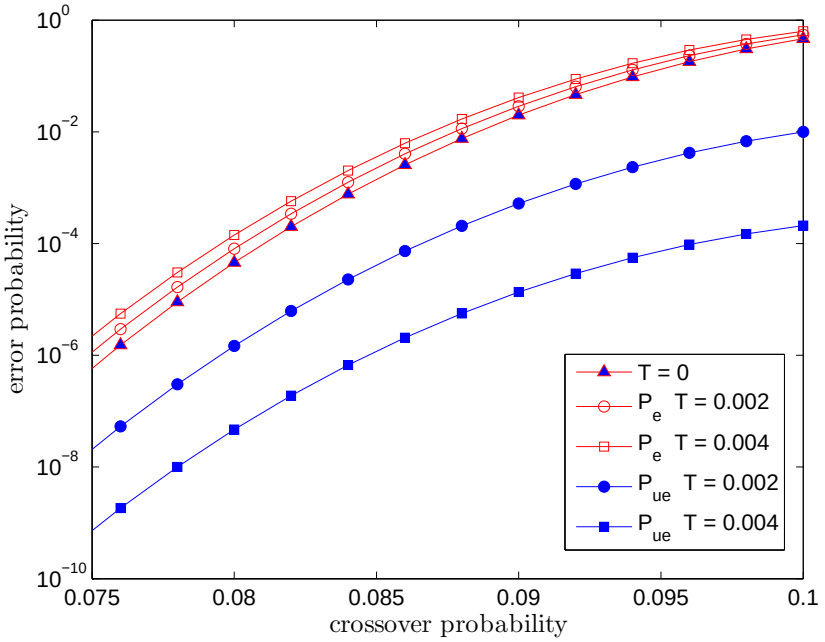
The fraction of the removed codebooks is upper bounded by ϵ_n. In the following examples, the value of ϵ_n is negligible. For the (6,12) regular binary ensemble with block lengths of $n = 504$ and 2004 bits, $\epsilon_n = 3.6002 \cdot 10^{-5}$, and $5.5058 \cdot 10^{-8}$, for $D_n = 40$ and 160 bits, respectively. For the (8,16) regular octal alphabet ensemble with a block length of $n = 1008$ symbols and $D_n = 80$ symbols, ϵ_n is around 10^{-14}.

2. *Performance over the AWGN channel*: For the AWGN channel, the results in this chapter are provided as function of the signal-to-noise ratio $\frac{E_s}{N_0}$ where E_s is the energy per transmitted coded symbols, and $\frac{N_0}{2}$ is the two-sided power spectral density of the additive white noise. This comment concerns both binary and non-binary codes.

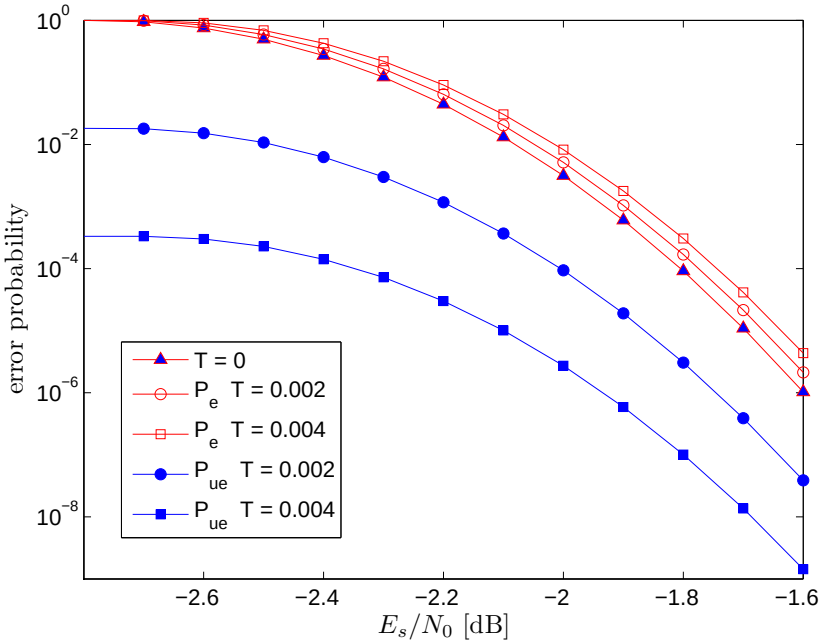
Example 3.3 (Error performance of binary regular LDPC code ensembles under generalized decoding with erasures) Consider an expurgation of the binary and regular (6,12) LDPC code ensemble of Gallager [44] with a block length of $n = 2004$ bits. In this expurgated ensemble, all the codebooks whose minimum distance is not larger than $D_n = 160$ are removed. Upper bounds on the block error probability and the undetected error probability, under Forney's generalized decoding with erasures, are studied based on Corollary 3.3. The composition spectrum is upper bounded via (3.32) and (3.33), where the composition spectrum of the original (non-expurgated) regular LDPC code ensemble is evaluated using the method provided in [18], [105]. The bounds are provided for several non-negative values of T in Figures. 3.3(a) and 3.3(b), assuming that the transmission takes place over a BSC and a binary-input AWGN channel, respectively. Note that if $T = 0$, the resulting bounds on the block error probability and the undetected error probability coincide, and they also provide an upper bound on the ML decoding error probability. The results indicate that by allowing an error probability that may be slightly higher than the upper bound on the error probability under ML decoding, significant improvement is guaranteed for the undetected error probability. Consider for example the error performance where the transmission takes place over a BSC with a crossover probability of 0.088. The upper bound on the error probability under ML decoding is around $7.5 \cdot 10^{-3}$ (see Figures. 3.3(a)). By allowing the total error probability to be less than $2 \cdot 10^{-2}$, the undetected errors are guaranteed to be less than $2 \cdot 10^{-4}$ and $5 \cdot 10^{-6}$ for $T = 0.002$ and 0.004, respectively.

Example 3.4 (Error performance of binary regular LDPC code ensembles under generalized decoding with a variable-size list) The performance of the same expurgated ensemble as in Example 3.3 is studied here under Forney's generalized decoding with a variable list-size. Upper bounds on the block error probability and the expected number of incorrect codewords in the list, are evaluated based on the bounds in Corollary 3.3 for several non-positive values of T. These bounds are provided in Figures. 3.4(a), and 3.4(b), assuming a transmission over a BSC or a binary-input AWGN channels, respectively. It is evident that only a slight improvement in the error performance is possible by using the generalized decoding rule. Take for example the case of transmission over a BSC: for crossover probabilities where the block error probability under ML decoding is below 0.09, the expected number of incorrect codewords is low. In fact, the upper bound on the expected number of incorrect codewords for such crossover probabilities, is less than one which implies that the list is likely to include only the correct codeword. However, for crossover probabilities for which the probability of the list error event is larger, the upper bound on the size of the decoded list grows considerably above 1 (see Figure. 3.4(a)).

Example 3.5 (Generalized decoding of non-binary regular LDPC code ensembles) Consider an expurgation of Gallager's ensemble of (8,16) regular LDPC codes [44] with an octal alphabet, and a block length of 1008 symbols. Consider the case where the expurgated ensemble excludes all the codebooks whose minimum distance is not larger than $D_n = 80$. The upper bounds on the error probabilities, under the generalized decoding rule in (3.4), are studied based on the upper bounds provided in Corollary 3.3. The (average) composition spectrum is upper bounded via (3.32) and (3.33), and the composition spectrum of the original ensemble is evaluated using the method provided in [34]. For the case of decoding with erasures, upper bounds on the block error and undetected error probabilities are provided, whereas for decoding with a variable list size, an upper bound on the expected number of incorrect codewords in the list and an upper bound on the block error probability are provided. These bounds are shown in Figures. 3.5(a) and 3.5(b), assuming that the transmission takes place over an 8-ary discrete memoryless symmetric channel, and an AWGN channel with 8-PSK modulation, respectively. It is evident that the upper bound on the block error probability for the case of decoding with erasures, referring to $T = 0.01$ in Figures. 3.5(a) and 3.5(b), slightly deteriorates as compared to the block error probability under ML decoding (where the bound presented for $T = 0$ coincides with the bound under ML decoding). However, a remarkable improvement is shown in these figures with resect to the undetected error probability (referring to

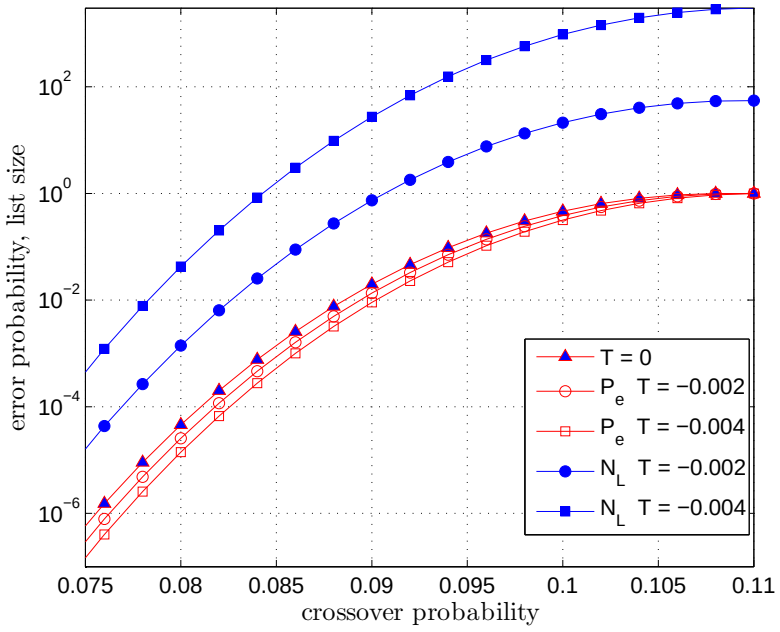


(a) Transmission over a BSC

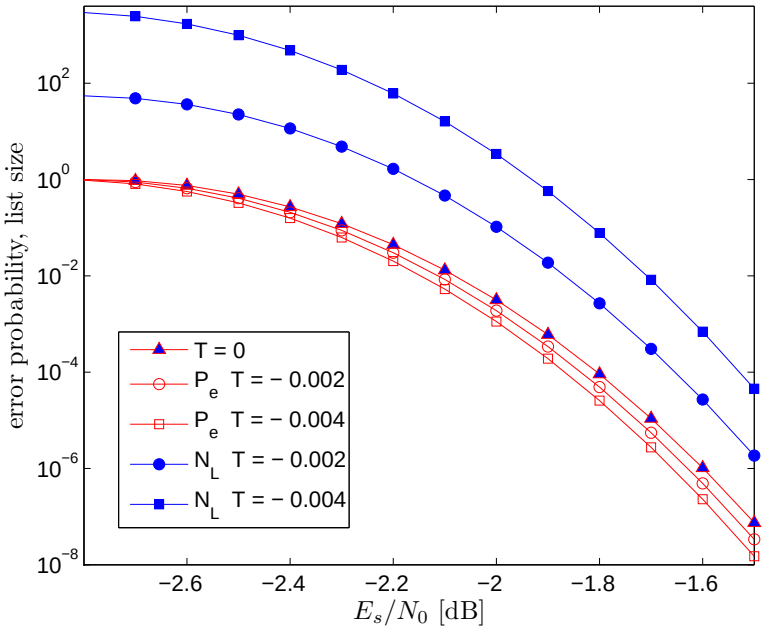


(b) Transmission over a binary-input AWGN channel

Figure 3.3: Upper bounds on the block error and undetected block error probabilities under the generalized decoding rule in (3.4) with erasures ($T \geq 0$). An expurgation of the binary and regular (6,12) LDPC code ensemble of Gallager is considered, where the block length is 2004 bits, and the parameter D_n which refers to the expurgation is set to 160 (see Example 3.3). The transmission in plots (a) and (b) is assumed to take place over a BSC, and a binary-input AWGN channels, respectively.



(a) Transmission over a BSC



(b) Transmission over a binary-input AWGN channel

Figure 3.4: Upper bounds on the block error probability and expected size of incorrect codewords in the decoded list, under the generalized decoding rule in (3.4) with variable-size list ($T \leq 0$). An expurgation of the binary and regular (6,12) LDPC code ensemble of Gallager is considered, where the block length is 2004 bits, and the parameter D_n which refers to the expurgation is set to 160 (see Example 3.3). The transmission in plots (a) and (b) is assumed to take place over a BSC, and a binary-input AWGN channels, respectively.

P_{ue} for $T = 0.01$ in both figures). For the variable-size list decoding which refers to $T = 0.01$ in (3.4), only a slight improvement is provided in the probability of error.

3.3 Applications to performance analysis of hybrid-ARQ systems

3.3.1 Preliminaries

Coded communication systems with one-bit noiseless feedback are considered where a generalized decoding rule with erasures is applied at the receiver. Each decoding erasure is communicated via the feedback to the transmitter, which then retransmits its message. It is first assumed that each transmitted block is decoded separately. Such a hybrid-ARQ system is described and studied in [41], where the error exponents for random coding are provided. For the case where deadlines are assumed, the error exponents for random coding are provided in [48].

The following discussion is provided in [41] and [48], and it is surveyed here for the sake of completeness.

Since Forney's generalized decoding rule (3.4) with a positive value of T is used in the context of erasures, the resulting decision regions at the receiver are disjoint, and the erasure probability P_{x} for a single block transmission is given by

$$P_{\mathrm{x}} = P_{\mathrm{e}} - P_{\mathrm{ue}}$$

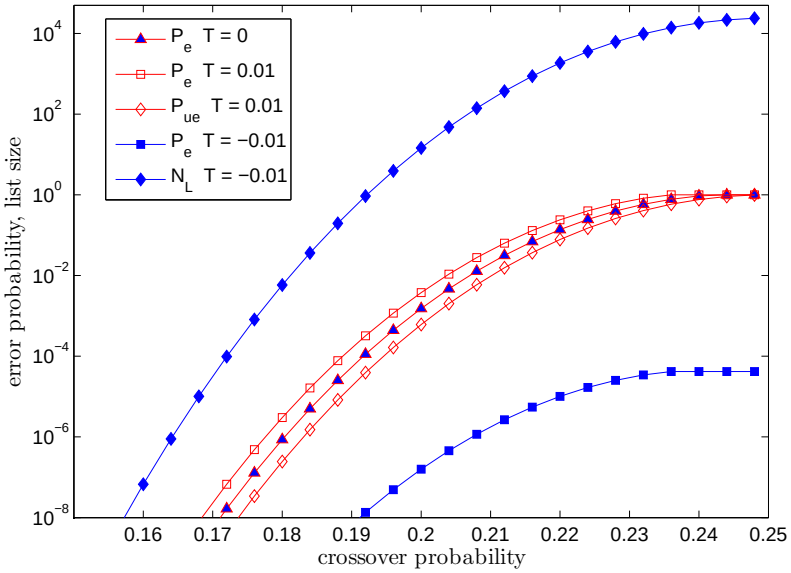
where P_{e} and P_{ue} are, respectively, the (total) block error probability and undetected error probability for a single block transmission. The erasure probability is studied via an upper bound on the error probability P_{e}. Assuming a noiseless and immediate feedback, for the case where no deadlines are considered, the expected rate of the considered system equals

$$(1 - P_{\mathrm{x}})R \tag{3.34}$$

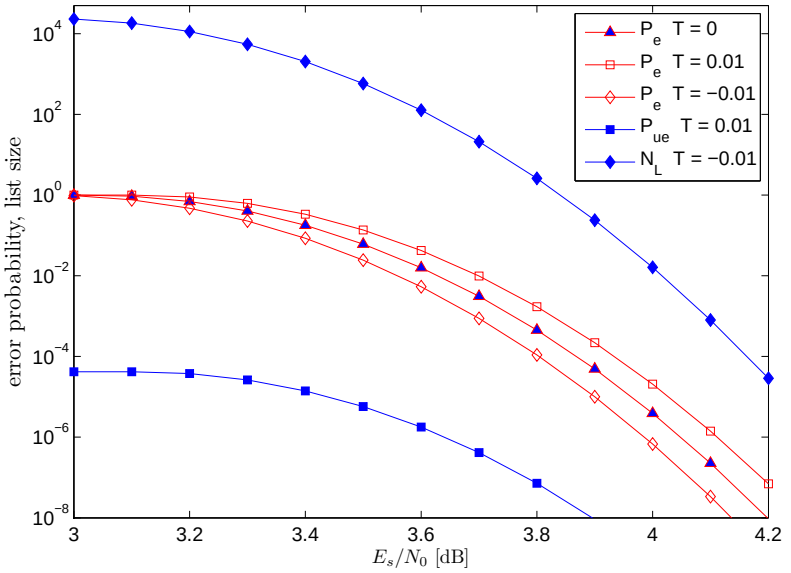
where R is the rate of the codebook used (in units of bits per channel use) for a single block transmission. The error probability of this scheme is given by

$$\frac{P_{\mathrm{ue}}}{1 - P_{\mathrm{x}}}. \tag{3.35}$$

Note that the replacement of P_{x} in (3.34) and (3.35) with an upper bound on P_{e}, provides a lower bound on the expected rate and an upper bound on the error probability.



(a) Transmission over an 8-ary discrete memoryless symmetric channel



(b) Transmission over an AWGN with 8-PSK modulation

Figure 3.5: Upper bounds on the decoding error probabilities and number of incorrect codewords in the decoded list for an expurgated ensemble of LDPC codes. The considered ensemble refers to the octal-alphabet regular (8,16) LDPC code ensemble of Gallager with a block length of 1008 symbols, and where the parameter D_n which refers to the expurgation is set to 80 (see Example 3.5). The upper bounds in Corollary 3.3 are provided in plots (a) and (b), assuming that the transmission takes place over an 8-ary discrete memoryless symmetric channel, and an AWGN channel with 8-ary PSK modulation, respectively.

For the case where deadlines are considered, let Q ($Q \geq 1$) be the maximal number of block retransmissions (including the first transmitted block). Each transmitted block is decoded separately using Forney's generalized decoding rule with erasures. Such a scheme is termed memoryless in [48] (note that the ARQ scheme without deadlines, studied in [41], is also memoryless in this sense). In cases where Q consequent block transmissions occur, then the generalized decoding rule is replaced for the last (Q-th) retransmitted block with an ML decoder. As a result, the expected rate and error probability, denoted by $R(Q)$ and $P_{\mathrm{e}}(Q)$, respectively, satisfy

$$\begin{aligned} R(Q) &= \frac{R}{\sum_{k=0}^{Q-1} (P_{\mathrm{x}})^k} \\ &= \frac{R\,(1-P_{\mathrm{x}})}{1-(P_{\mathrm{x}})^Q} \end{aligned} \tag{3.36}$$

and

$$\begin{aligned} P_{\mathrm{e}}(Q) &= \sum_{k=1}^{Q-1} (P_{\mathrm{x}})^{k-1}\, P_{\mathrm{ue}} + (P_{\mathrm{x}})^{Q-1}\, P_{\mathrm{e}}^{\mathrm{ML}} \\ &= \frac{\left(1-(P_{\mathrm{x}})^{Q-1}\right)\, P_{\mathrm{ue}}}{1-P_{\mathrm{x}}} + (P_{\mathrm{x}})^{Q-1}\, P_{\mathrm{e}}^{\mathrm{ML}} \end{aligned} \tag{3.37}$$

where $P_{\mathrm{e}}^{\mathrm{ML}}$ is the block error probability under ML decoding for the considered code (while referring to the decoding of the last retransmitted block separately). Note that in the limit where $Q \to \infty$ (no deadlines), then (3.36) and (3.37) tend asymptotically to (3.34) and (3.35), respectively. Replacing P_{x} in (3.36) and (3.37) with an upper bound on the (total) error probability P_{e}, results in a lower bound on the expected rate, and an upper bound on the error probability, respectively.

In hybrid incremental-redundancy ARQ schemes, a repeat request triggers the transmission of a new block of n coded symbols which is not necessarily equal to the former block (even though the transmission of the same message is concerned). The decoder, instead of processing only the last block, decodes the message by observing the entire blocks received so far for the concerned message. For such cases, the expected rate, denoted by $R^{\mathrm{IR}}(Q)$, satisfies the following lower bound [48, Eq. (24)]:

$$R^{\mathrm{IR}}(Q) \geq \frac{R}{1+(Q-1)P_{\mathrm{x}}}. \tag{3.38}$$

This bound coincides with (3.36) if $Q = 2$. However, for $Q > 2$, the bound in (3.38) is loosened because of the specific derivation used in [48]. Assuming that an ML decoder is used after the last retransmitted block, the error probability for the IR-ARQ scheme, denoted by $P_{\mathrm{e}}^{\mathrm{IR}}(Q)$, is upper bounded by [48, Eq. (25)]:

$$P_{\mathrm{e}}^{\mathrm{IR}}(Q) \leq \sum_{k=1}^{Q-1} P_{\mathrm{ue}}(k) + P_{\mathrm{e}}^{\mathrm{ML}}(Q) \tag{3.39}$$

where $P_{\text{ue}}(k)$ denotes the undetected error probability of the generalized decoding rule, which operates on the received observations of k consequent transmitted blocks ($1 \leq k \leq Q-1$), and $P_{\text{e}}^{\text{ML}}(Q)$ denotes the error probability under ML decoding, based on the entire transmission of Q blocks (the ML decoder is used only if Q blocks are needed to be transmitted for the same message). Note that the dominant summand in (3.39) is $P_{\text{ue}}(1)$, i.e., the undetected error probability of the first transmitted block.

3.3.2 Examples

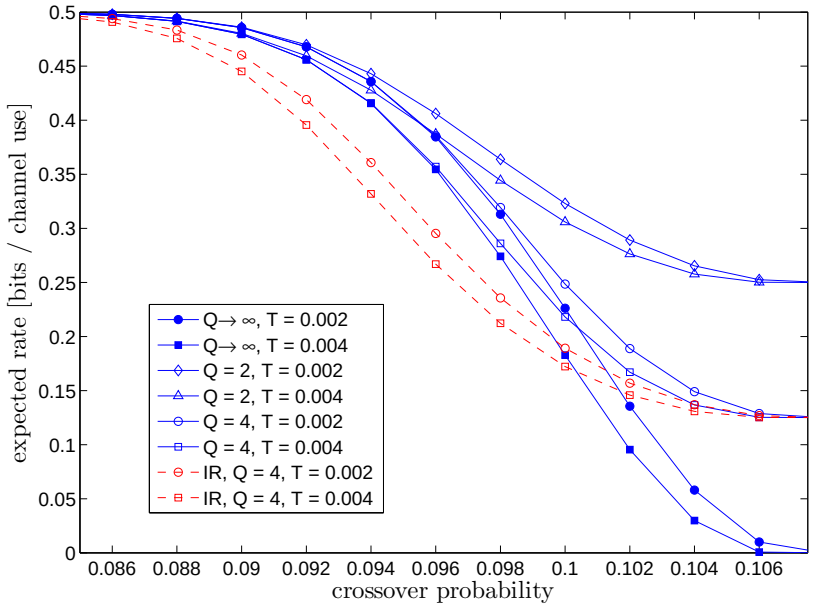
In the following examples, upper bounds on the error performance and lower bounds on the expected rates of some hybrid-ARQ systems are studied. These bounds are based on the bounds in Corollary 3.3 and the results in Section 3.3.1. As mentioned, each block of coded symbols in the IR-ARQ scheme may include new coded symbols. Nevertheless, for all examples in this section where IR-ARQ schemes are considered, a retransmission of equal coded blocks is assumed.

Example 3.6 (Hybrid-ARQ schemes over BSC) Consider the expurgated ensemble of binary regular LDPC codes in Example 3.3, whose transmission takes place over a BSC. Lower bounds on the expected rates are presented for several values of the decoding parameter T in Figure 3.6(a). For memoryless systems without deadlines, the provided lower bound on the expected rate in (3.34) drops to zero as the crossover probability of the BSC approaches the capacity limit (which is 0.11 for a design rate of $R = \frac{1}{2}$ bits per channel use). For schemes with deadlines of $Q = 2$ and 4 transmissions, the lower bounds on the expected rate in (3.36) drop to $\frac{R}{Q} = \frac{1}{4}$ and $\frac{1}{8}$, respectively, as the crossover probability of the BSC approaches the capacity limit (which is the limit of (3.36) when we let P_{x} tend to 1). Schemes with incremental redundancy are also considered. Note that the lower bound on the expected rates for memoryless schemes with deadline of $Q = 2$, also applies to schemes with incremental redundancy, the lower bound in (3.38) coincides with the equality in (3.36) for $Q = 2$. For the case of $Q = 4$, the loosened lower bound on the expected rate for incremental redundancy schemes in (3.38) is also provided. Upper bounds on the decoding error probabilities for the considered schemes are provided in Figure 3.6(b). The upper bound for a block error probability with $T = 0$ and where no feedback is available (a single transmission, $Q = 1$) is also provided. Note that this bound is valid for the block error probability under ML decoding. Comparing this upper bound (for $T = 0$ and $Q = 1$), with the upper bounds for $T = 0.002$ and 0.004, shows that the introduction of one-bit immediate and noiseless feedback allows for a considerable

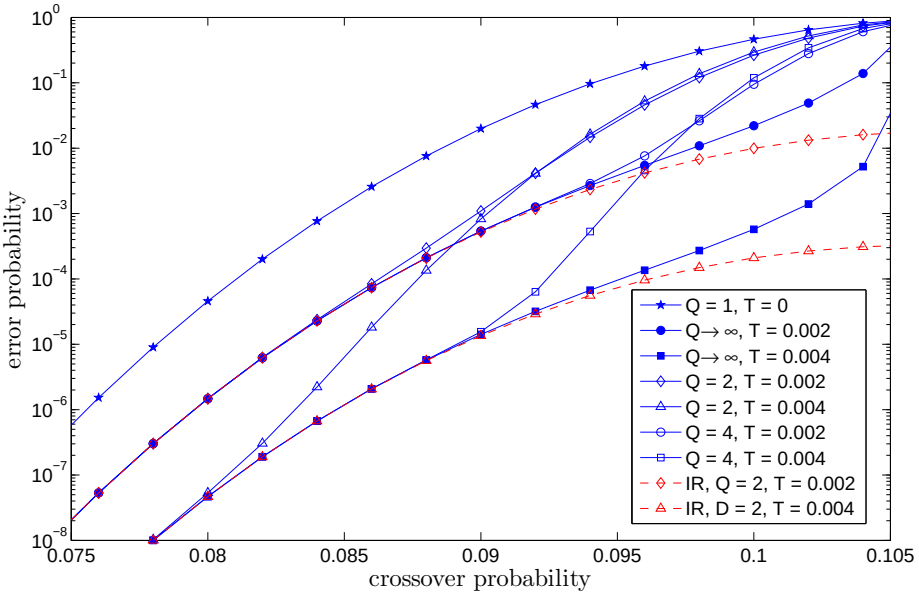
improvements in the error performance. This improvement is achieved while maintaining reasonable rate drops (at least for crossover probabilities below the threshold for which the rate starts dropping considerably). Moreover, the improvement is of interest even for the simplified memoryless-ARQ schemes with moderate deadlines (of $Q = 2$ and 4 block transmissions).

Example 3.7 (Hybrid-ARQ schemes over binary-input AWGN channels) Consider the expurgated, binary, and regular LDPC code ensemble in Example 3.3, and the hybrid-ARQ scheme used in Example 3.6. Lower bounds on the expected rates, and upper bounds on the error probabilities for such schemes are provided in Figures. 3.7(a), and 3.7(b), respectively, assuming that transmission takes place over a binary-input AWGN channel. The results show that if the SNR is above a threshold for which the expected rate does not deteriorate considerably, a substantial improvement in the decoding error probability is possible. This improvement is achieved while maintaining a negligible rate loss, even for the simplified memoryless schemes with moderate deadlines (e.g., $Q = 2$ and 4). Take for example the case where $E_s/N_0 = -2.1$ dB. For this setting, the upper bound on the error probability under ML decoding without retransmissions ($T = 0$, $Q = 1$) is slightly above 10^{-2}. By introducing a one bit noiseless feedback, the upper bounds on the error probability for all considered schemes with $T = 0.004$ are in the range of $10^{-4} - -10^{-5}$ while maintaining a small rate loss (the rate loss for the memoryless scheme with deadlines of $Q = 2$ transmissions is below 3.2%).

Example 3.8 (Hybrid-ARQ schemes over AWGN channels with non-binary LDPC codes) Hybrid ARQ schemes over the AWGN channel with 8-PSK modulation is considered where the expurgated and octal-alphabet LDPC code ensemble in Example 3.5 is used. Lower bounds on the expected rate and upper bounds on the decoding error probability are shown in Figures. 3.8(a) and 3.8(b), respectively. Schemes with and without deadlines are considered. The results show that the lower bounds on the expected rates drop considerably, below $E_s/N_0 = 3.6$ dB. However, above this SNR, the introduction of a single-bit, noiseless and immediate feedback allows to achieve remarkable improvements in the error performance. Take for example the case where $E_s/N_0 = 3.62$ dB where the upper bound on the error probability under ML decoding without feedback (see the curve for $T = 0$ and $Q = 1$) is around 10^{-2}. For the same channel, if no deadlines are assumed, the upper bounds on the error probability are around $2 \cdot 10^{-6}$. When deadlines of $Q = 2$ and 4 total retransmissions (including the first transmission) are assumed, the upper bounds on the

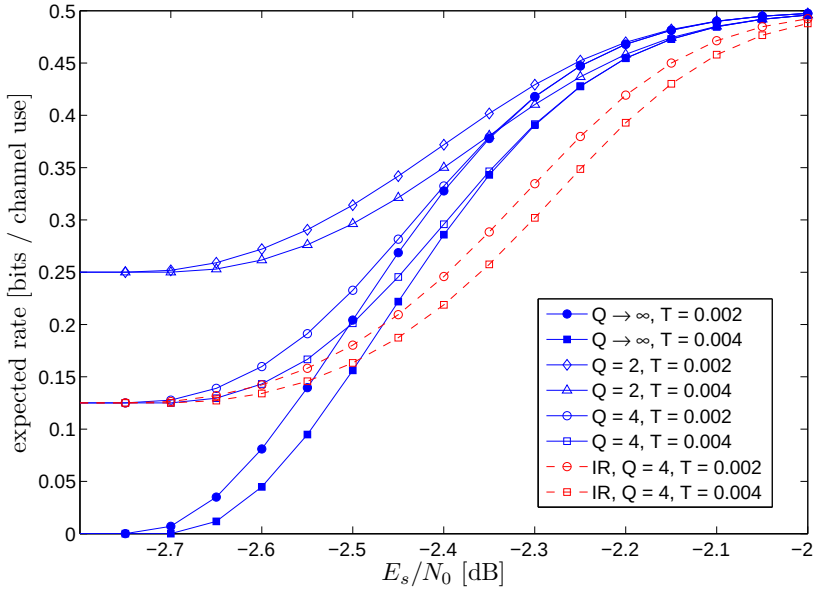


(a) Lower bounds on the expected rates

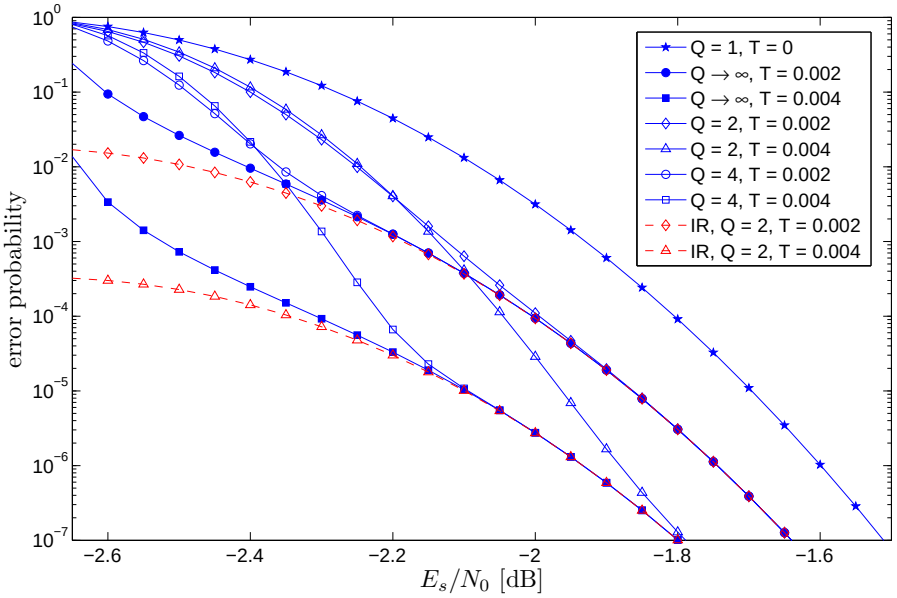


(b) Upper bounds on the error probability

Figure 3.6: Performance bounds of hybrid-ARQ schemes for the expurgated, binary and regular (6,12) LDPC code ensemble of Gallager with a block length of $n = 2004$ bits (see Example 3.3). The transmissions are assumed to take place over the BSC. In plot (a), lower bounds on the expected rates for memoryless hybrid-ARQ schemes with and without deadlines (see (3.36), and (3.34), respectively) are shown for $T = 0.002$ and 0.004 (and deadlines of $Q = 2$ and 4 transmissions). In plot (b), upper bounds on the error probability are provided for the considered schemes. For the case of $Q = 2$, lower bounds on the expected rate and upper bounds on the decoding error probability are also provided in plots (a) and (b), respectively, assuming incremental-redundancy ARQ at the decoder (see (3.38)).

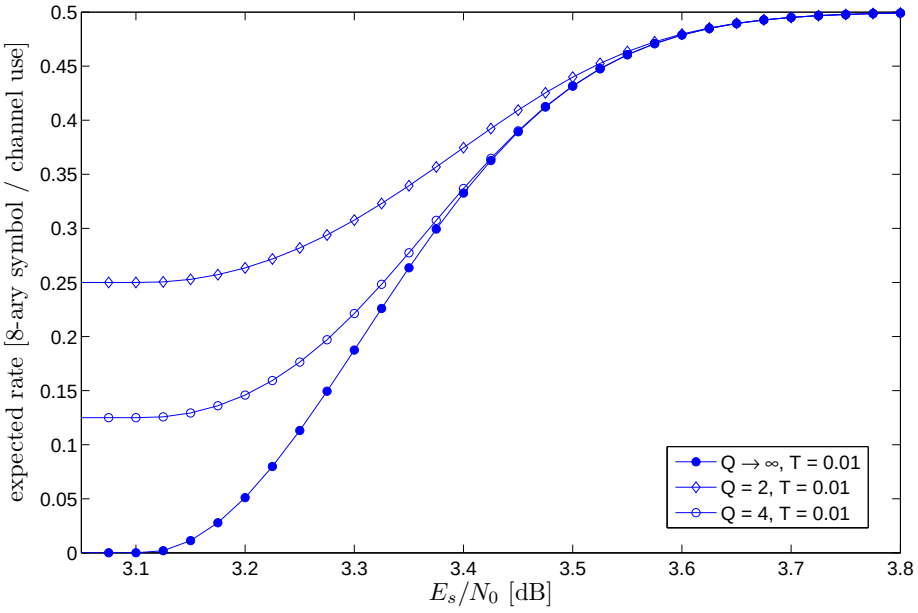


(a) Lower bounds on the expected rates

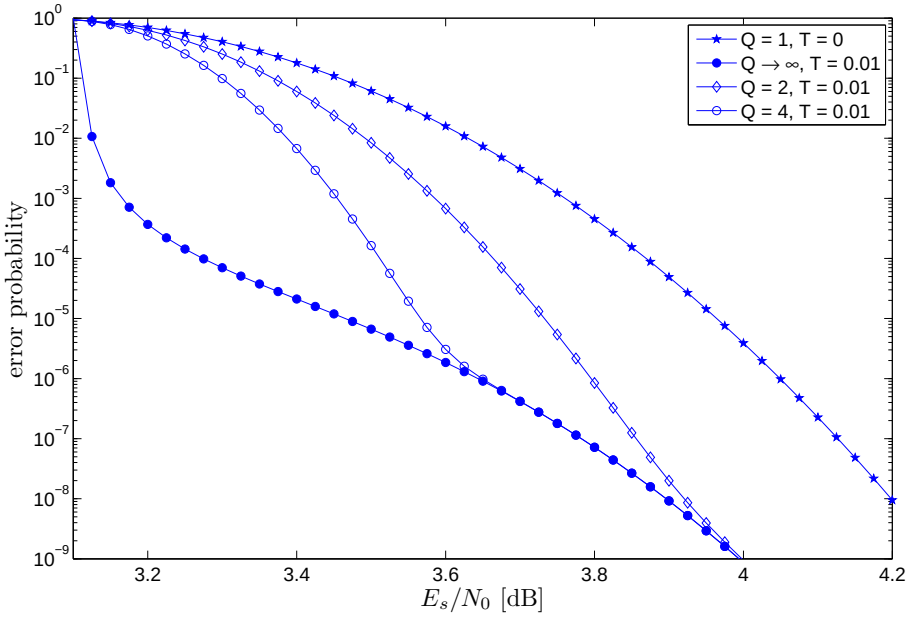


(b) Upper bounds on the error probability

Figure 3.7: Performance bounds of hybrid-ARQ schemes for the expurgated, binary and regular (6,12) LDPC code ensemble of Gallager with a block length of $n = 2004$ bits (see Example 3.3). The transmissions are assumed to take place over binary-input AWGN channels. In plot (a), lower bounds on the expected rates for memoryless hybrid-ARQ schemes with and without deadlines (see (3.36), and (3.34), respectively) are shown for $T = 0.002$ and 0.004 (and deadlines of $Q = 2$ and 4 transmissions). In plot (b), upper bounds on the error probability are provided for the considered schemes. For the case of $Q = 2$, the lower bounds on the expected rate and upper bounds on the decoding error probability are also provided in plots (a) and (b), respectively, assuming incremental-redundancy ARQ at the decoder (see (3.38)).



(a) Lower bounds on the expected rates



(b) Upper bounds on the error probability

Figure 3.8: Performance bounds of hybrid-ARQ schemes based on an expurgated, octal-alphabet and regular (8,16) LDPC code ensemble with a block length of $n = 1008$ symbols (see Example 3.5). The transmission is assumed to take place over an AWGN channel with 8-PSK modulation. In plot (a), lower bounds on the expected rates for memoryless hybrid-ARQ schemes with and without deadlines (see (3.36), and (3.34), respectively) are shown for $T = 0.01$ (and possible deadlines of $Q = 2$ and 4 transmissions). In plot (b) upper bounds on the error probability are provided for the considered schemes.

error probability for the same channel are $6 \cdot 10^{-4}$ and $3 \cdot 10^{-6}$, respectively. For all considered schemes, the expected rate deteriorates at this point by no more than 4%.

Immediate and noiseless one-bit feedback is assumed in Examples 3.6-3.8. The restriction to immediate feedback is loosened in most network applications where some sort of a multiple-access protocol is introduced. As a result of the applied protocol, the transmitter is informed regarding the one-bit feedback with some delay that is guaranteed (by the protocol) to be before the next time slot of the retransmission. As for the condition of noiseless feedback, loosening this condition results in an inevitable synchronization errors (see, e.g., a similar observation in [30]). Since the hybrid-ARQ schemes presented in this section require only one-bit feedback, even if these synchronization errors should be kept low in comparison with the block error performance, they are typically achievable with relatively low resources.

3.4 Upper Bounds under suboptimal decoding with erasures

In this section, upper bounds on decoding error probabilities are derived for the suboptimal decoding rule in (3.6).

Proposition 3.7 Consider the transmission of a block code $\mathcal{C}$ of block length n and M codewords, and let $p(\mathbf{y}|\mathbf{x})$ designate the transition probability of the channel where $\mathbf{x} \in \mathcal{C}$ is the transmitted codeword and $\mathbf{y} \in \mathcal{Y}^n$ is the received vector. Then, the conditional block error probability $P_{\mathrm{e}|m}$, and the conditional undetected error probability $P_{\mathrm{ue}|m}$, under the suboptimal decoding rule in (3.6) satisfy

$$P_{\mathrm{e}|m} \leq e^{nsT} D_{\mathrm{B}}(m, G_n^m, s, \rho), \ 0 \leq s \leq \rho \leq 1 \tag{3.40}$$

$$P_{\mathrm{ue}|m} \leq e^{-nsT} D_{\mathrm{B}}(m, G_n^m, s, \rho), \ 0 \leq s \leq \rho \leq 1 \tag{3.41}$$

where $D_{\mathrm{B}}(m, G_n^m, s, \rho)$ is defined in (3.10), and G_n^m is an arbitrary non-negative function over $\mathcal{Y}^n$ which possibly depends on the codeword $\mathbf{x}_m$, $1 \leq m \leq M$.

Proof: See Appendix 3.G. ∎

Remark 3.12 The upper bound on the block error probability in (3.40) coincides with the upper bound on the total error probability provided in (3.8) under the optimal generalized decoding rule. On the other hand, the upper bounds on the undetected error probabilities under the optimal and suboptimal decoding rules in (3.9) and (3.41), respectively, are different.

The following corollary is a particularization of Proposition 3.7 for the ensemble of fully random block codes of length n and rate R whose transmission takes place over memoryless channels:

Corollary 3.4 Consider the transmission of block codes over a memoryless communication channel. Then, there exists a block code satisfying

$$P_{\mathrm{e}} \leq e^{-nE_1(R,T)}$$
$$P_{\mathrm{ue}} \leq e^{-nE_2^*(R,T)}$$

where $R \triangleq \frac{\ln M}{n}$ is the code rate (in nats per channel use), $E_1(R,T)$ is defined in (3.13),

$$E_2^*(R,T) \triangleq \max_{0 \leq s \leq \rho \leq 1,\ q_X} \Big(E_0(s,\rho,q_X) - \rho R + sT \Big)$$

E_0 is as defined in (3.14), and q_X is an arbitrary probability distribution over $\mathcal{X}$.

Proof: The proof follows the same arguments as the proof of Corollary 3.1. ■

The following bound is provided for the case of binary linear block codes whose transmission takes place over an MBIOS channel (the generalization of the bound to non-binary linear block codes, as provided in Chapter 2, is direct):

Corollary 3.5 Consider an (n,k) binary linear block code $\mathcal{C}$ whose transmission takes place over an MBIOS channel with a transition probability law p. Then the block error probability P_{e}, and the undetected error probability P_{ue}, under the generalized decoding rule in (3.6) satisfy

$$P_{\mathrm{e}} \leq e^{-n\left(E(\rho,R,\mathcal{C}) - \frac{\rho T}{1+\rho}\right)},\ 0 \leq \rho \leq 1 \tag{3.42}$$
$$P_{\mathrm{ue}} \leq e^{-n\left(E(\rho,R,\mathcal{C}) + \frac{\rho T}{1+\rho}\right)},\ 0 \leq \rho \leq 1 \tag{3.43}$$

where R is the code rate (in nats per channel use), and $E(\rho,R,\mathcal{C})$ is defined in (3.20).

Proof: The proof follows from Proposition 3.7, and its derivation is similar to the way where Corollary 3.2 is derived from Proposition 3.6. ■

Remark 3.13 As in Corollary 3.2, the bounds of Corollary 3.5 resemble to the SFB, and they may therefore be considered as a generalization of the SFB for the case at hand.

Remark 3.14 For all rates below some (finite) rate thresholds, the bounds in Corollary 3.5 on the decoding error for linear block codes under the suboptimal LR rule in Definition 3.2, coincide with those under the optimal decoding rule in Definition 3.1. To see this, observe first that the upper bounds in (3.18) and (3.42) are identical. It is left to consider the upper bounds in (3.19) and (3.43) on the undetected error probability. Note first that $E_0(\rho) - \rho R$ (E_0 is defined in (3.21)) is a concave function of $0 \leq \rho \leq 1$, and it is optimized for rates below $E_0'(1)$ at $\rho = 1$ (see, e.g., [111, p. 135]). Moreover, $\frac{\rho}{1+\rho}$ is a monotonic increasing function of $0 \leq \rho \leq 1$. This implies that if $\frac{T}{4} < E_0'(1)$, then at all rates below $E_0'(1) - \frac{\ln(\alpha(\mathcal{C}))}{n} - \frac{T}{4}$, the error exponents of the upper bounds in (3.19) and (3.43) are both maximized at $\rho = 1$, and they therefore coincide. A similar observation is provided in [54, p. 82] for the ensemble of fully random block codes. Specifically, it is observed in [54] that up to some rate threshold, the upper bounds under the suboptimal LR decoding rule for the ensemble of fully-random block codes coincide exponentially with those provided by Forney in [41].

Example 3.9 (Error exponents of fully random binary linear block codes) Fully random binary and linear (n, k) block codes are considered where, as mentioned in Example 3.2, $\alpha(\mathcal{C}) = 1$ (see (3.22)). For the particular case of transmission over a BSC, the error exponents for the considered ensemble are studied in [9] and [16]. The lower bounds on the block error exponents and the undetected error exponents from [9] and [16] are compared in Figures 3.9(a), and 3.9(b), respectively, to the bounds provided in Corollary 3.5. The bounds are derived for a BSC with a crossover probability of $p = 0.07$ and a decoding parameter $\tau = 0.03$ (see (3.7) where these are the same parameters studied in [9, Figure 1]). The error exponent provided by Gallager for the case of ML decoding is also provided for comparison, in addition to the undetected error exponent under the optimal generalized decoding rule. Apart from low rates, where the bounds in [9] and [16] outperform those provided in Corollary 3.5, the latter bounds on the error exponents lie in between the two previously reported bounds from [9] and [16] (see Figure 3.9). Moreover, in the rate region beyond the critical rate, where the bound in [9] outperform the bound in [16], the derived bounds perform in close proximity to the tightest known bound. The superiority of the undetected error exponent under the optimal decoding rule is clearly pronounced. This comparison is further studied in Figure. 3.10 where the lower bounds on the undetected error exponents under the optimal and suboptimal generalized-decoding rules are provided for the same parameters as in Example 3.2 ($T = 0$, 0.025, 0.05, 0.1 and 0.15), assuming that transmission takes place over a BSC with a crossover probability of $p = 0.11$, and over binary-input AWGN channel with $E_s/N_0 = -2.8$ dB.

For the case where $T = 0$, both considered exponents, for optimal and suboptimal generalized-decoding rules, coincide with each other and with the (non-expurgated) random coding error exponent of Gallager [45]. As observed in Remark 3.14, it is evident that for low to moderate code rates, the bounds under optimal and suboptimal generalized decoding rules coincide. However, as the coding rates approach the channel capacity, the lower bounds on the undetected block error exponents under the suboptimal generalized-decoding, are considerably loosened in comparison to the lower bound under the optimal generalized decoding.

Corollary 3.6 Under the assumptions and notation in Corollary 3.3, the block error probability P_{e} and the undetected error probability P_{ue} under the suboptimal decoding rule in (3.6), satisfy

$$P_{\mathrm{e}} \leq e^{\frac{n\rho T}{1+\rho}} \cdot D_{\mathrm{s}}(\rho, \mathcal{C}), \quad 0 \leq \rho \leq 1 \tag{3.44}$$

$$P_{\mathrm{ue}} \leq e^{-\frac{n\rho T}{1+\rho}} \cdot D_{\mathrm{s}}(\rho, \mathcal{C}), \quad 0 \leq \rho \leq 1 \tag{3.45}$$

where $D_{\mathrm{s}}(\rho, \mathcal{C})$ is defined in (3.28).

Proof: Setting $s = \frac{\rho}{1+\rho}$, $G_n^m(\mathbf{y}) = \prod_{i=1}^n g(y_i)$ where g is as defined in (3.23), the proof follows from Proposition 3.7 in the same way as the proof of Theorem 2.3. ■

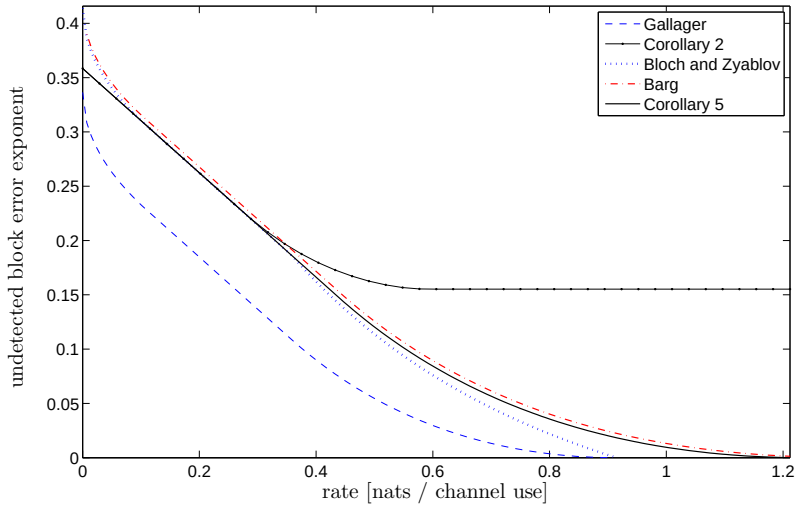
Consider the particular case of binary linear block codes whose transmission takes place over the binary-input AWGN channel with BPSK modulation. The bound of Divsalar (see [26] and [94, Sec. 3.2.4]) provides a closed-form expression for an upper bound on the block error probability under ML decoding. The following proposition provides a similar bound under the LR decoding rule in Definition 3.2:

Proposition 3.8 Consider the transmission of a binary linear block code over the AWGN channel with BPSK modulation, then the error and undetected error probabilities under the LR decoding in (3.6) satisfy

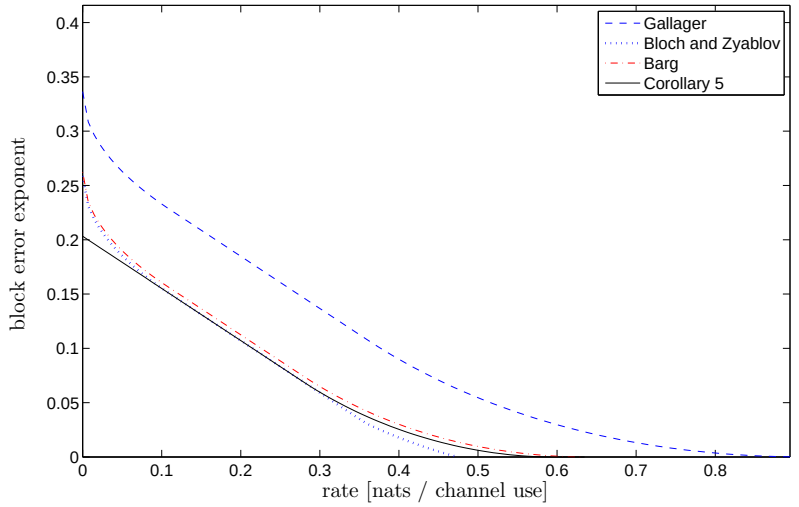
$$P_{\mathrm{e}} \leq \sum_{d=d_{\min}}^{n} \min \left\{ \exp\left(-nE_{\mathrm{e}}\left(\frac{d}{n}, \frac{E_{\mathrm{s}}}{N_0}\right)\right), |\mathcal{C}_d|\, Q\left(\sqrt{\frac{2E_{\mathrm{s}}d}{N_0}} - \frac{nT}{2\sqrt{\frac{2dE_{\mathrm{s}}}{N_0}}}\right) \right\} \tag{3.46}$$

$$P_{\mathrm{ue}} \leq \sum_{d=d_{\min}}^{n} \min \left\{ \exp\left(-nE_{\mathrm{ue}}\left(\frac{d}{n}, \frac{E_{\mathrm{s}}}{N_0}\right)\right), |\mathcal{C}_d|\, Q\left(\sqrt{\frac{2E_{\mathrm{s}}d}{N_0}} + \frac{nT}{2\sqrt{\frac{2dE_{\mathrm{s}}}{N_0}}}\right) \right\} \tag{3.47}$$

where $d_{\min}$ is the minimum Hamming distance of the code, n is the block length of the code, $|\mathcal{C}_i|$ is the number of codewords whose Hamming weight equals i, T is the

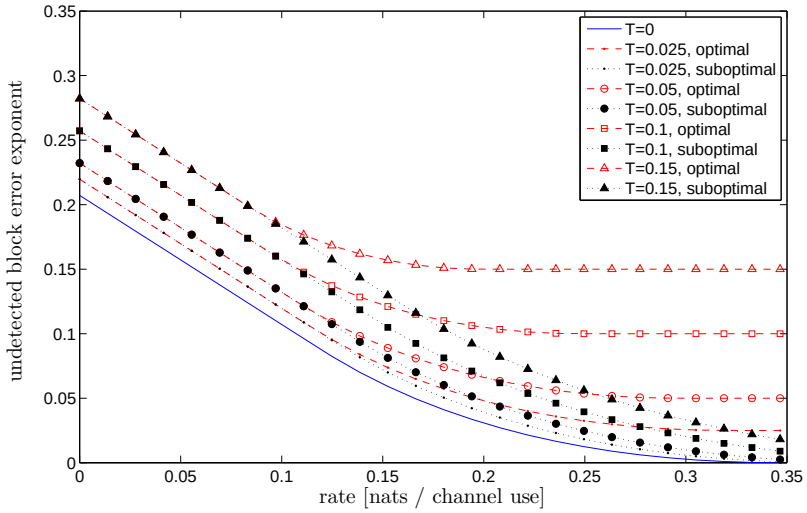


(a) Undetected error exponents

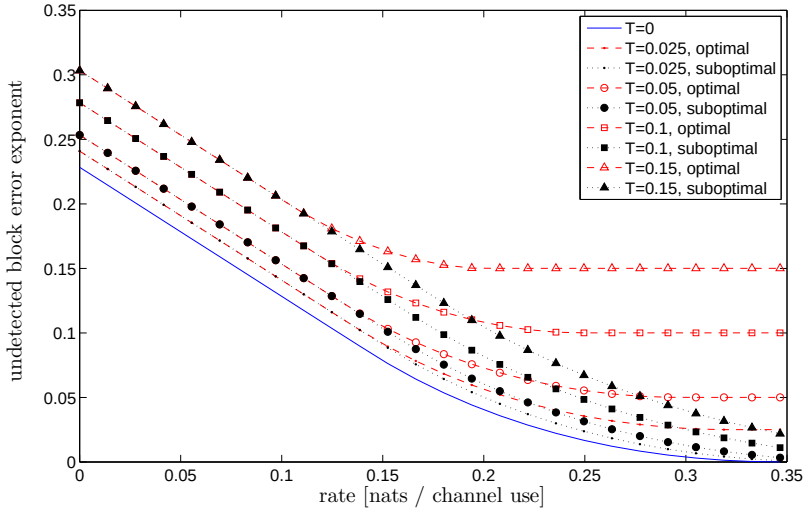


(b) Error exponents

Figure 3.9: Lower bounds on the block error exponents of fully-random binary linear block codes whose transmission takes place over a BSC with a crossover probability of $p = 0.07$, under the suboptimal decoding rule in (3.7) with $\tau = 0.03$. The lower bounds on the undetected block error exponents in [9, Theorem 2], [16] (see also [9, Theorem 1]), and Corollary 3.5 (see (3.43)) are provided in plot (a), together with Gallager's random-coding error exponent under ML decoding [45], and the lower bound on the undetected error exponent in Corollary 3.2 (see (3.19)) under the optimal generalized decoding rule. The lower bounds on the error exponents in [9, Theorem 2], [16], and Corollary 3.5 (see (3.42)) are provided in plot (b) (the lower bound of Gallager for the random-coding error exponent under ML decoding is also provided for comparison).



(a) Transmission over a BSC



(b) Transmission over a binary-input AWGN Channel

Figure 3.10: Lower bounds on the undetected error exponents of fully-random binary linear block codes under the suboptimal generalized decoding rule in (3.6). The bounds based on Corollary 3.5, are provided in plots (a) and (b), assuming that the transmission takes place over a BSC with a crossover probability of $p = 0.11$, and a binary-input AWGN channel with $E_s/N_0 = -2.8$ dB, respectively. The lower bounds on the error exponents under the optimum generalized decoding rule in (3.4), studied in Example 3.2, are also provided for comparison.

decoding parameter in (3.6), E_s is the energy per transmitted (coded) symbol, $\frac{N_0}{2}$ is

the two-sided power spectral density of the white Gaussian noise, and

$$E_{\mathrm{e}}\left(\delta, \frac{E_{\mathrm{s}}}{N_0}\right) \triangleq E_{\mathrm{D}}\left(\delta, \frac{E_{\mathrm{s}}}{N_0}\right) - \frac{T\xi}{2},$$

$$E_{\mathrm{ue}}\left(\delta, \frac{E_{\mathrm{s}}}{N_0}\right) \triangleq E_{\mathrm{D}}\left(\delta, \frac{E_{\mathrm{s}}}{N_0}\right) + \frac{T\xi}{2}$$

$$E_{\mathrm{D}}\left(\delta, \frac{E_{\mathrm{s}}}{N_0}\right) \triangleq -r_n(\delta) + \frac{1}{2}\ln\left(\beta + (1-\beta)e^{2r_n(\delta)}\right) + \frac{\beta\delta}{1-(1-\beta)\delta}\frac{E_{\mathrm{s}}}{N_0}$$

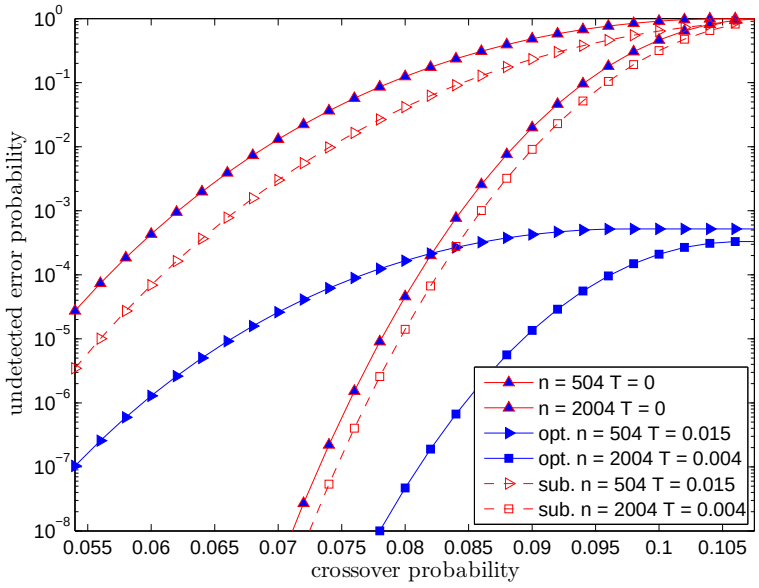
$$\beta \triangleq \sqrt{\frac{E_{\mathrm{s}}}{N_0}\frac{2(1-\delta)}{\delta(1-e^{-2r_n(\delta)})} + \left(\frac{1-\delta}{\delta}\right)^2\left(\left(1+\frac{E_{\mathrm{s}}}{N_0}\right)^2 - 1\right)} - \frac{1-\delta}{\delta}\left(1+\frac{E_{\mathrm{s}}}{N_0}\right)$$

$$r_n(\delta) \triangleq \frac{\ln|\mathcal{C}_d|}{n}, \quad \delta \triangleq \frac{d}{n}$$

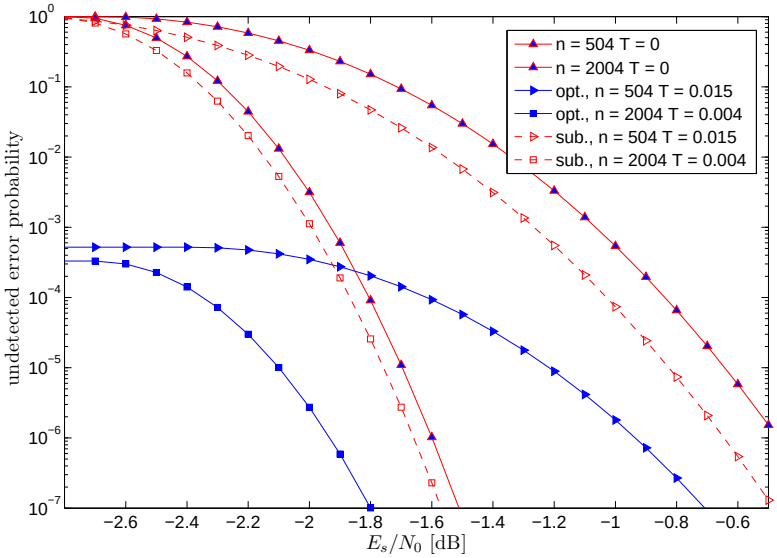
$$\xi \triangleq \frac{\beta}{\beta + (1-\beta)(1-\delta)}.$$

Proof: See Appendix 3.H. ∎

Example 3.10 (Error performance of expurgated binary and regular LDPC code ensembles under suboptimal generalized decoding with erasures) Consider an expurgation of the binary and regular LDPC code ensembles in Example 3.3 (with block lengths of 504 and 2004 bits). The upper bound in (3.45), on the undetected error probability under the generalized decoding rule with erasures in (3.6), is provided in Figures 3.11(a) and 3.11(b), assuming that the transmission takes place over a BSC and a binary-input AWGN channel, respectively. The upper bounds under the optimal generalized decoding rule are also provided for a comparison, in addition to the upper bound under the generalized decoding rule with $T = 0$ (which coincides with the upper bound on the error probability under ML decoding). It is evident that the resulting bounds under the suboptimal generalized decoding rule are loosened in comparison to the bounds under the optimal generalized decoding rule. This result is expected from the previous example where the undetected error exponents are studied for fully-random linear block codes. In Figure 3.12, the upper bounds on the undetected error probability in Corollary 3.6 are compared with those provided in Proposition 3.8. The provided bounds are for the binary regular and expurgated LDPC code ensembles in Example 3.3 (with block lengths of 504 and 2004 bits), and for a similar ensemble with a block length of 10008 bits and $D_n = 800$. The parameter T in (3.6) is chosen, for this comparison, to be 0.0198, 0.0050, and $9.992 \cdot 10^{-4}$, respective to the considered block lengths. It is evident that the simple bound in (3.47) is loosened in comparison to the bound in (3.45), but only by a relatively small difference.



(a) Transmission over a BSC



(b) Transmission over a binary-input AWGN channel

Figure 3.11: Upper bounds on the undetected error probabilities of some expurgated ensembles of binary and regular (6,12) LDPC codes under the optimal and sub-optimal generalized decoding rules in (3.4) and (3.6), respectively. The upper bound in Corollary 3.6 is shown in plots (a) and (b), assuming that the transmission takes place over a BSC and a binary-input AWGN channel, respectively. The upper bounds in Corollary 3.3, studied in Examples 3.3 and 3.4, are also provided for comparison.

3.5 Upper bounds under fixed-size list decoding

In this section, upper bounds on the block error probability are derived for the fixed-size list decoding (see Definition 3.3). As mentioned in Section 3.1, the block error

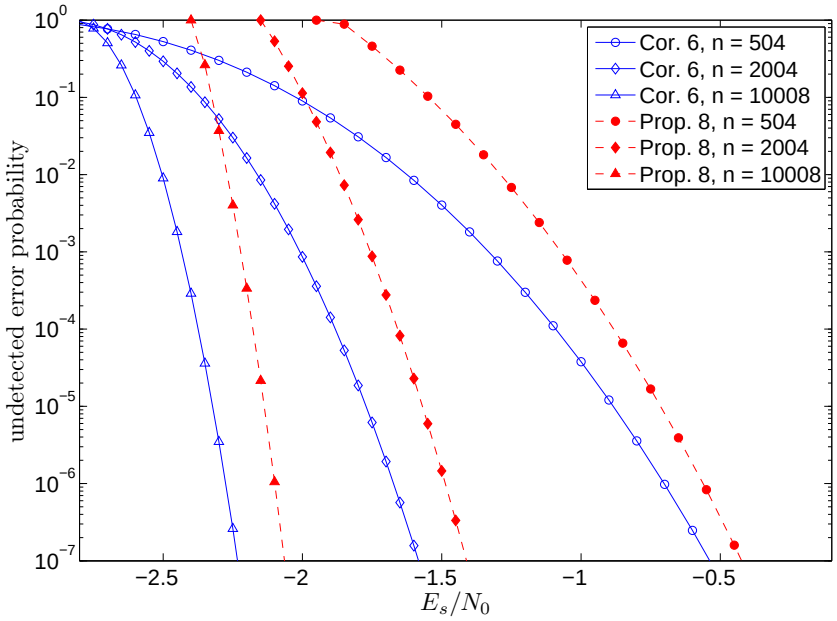


Figure 3.12: A comparison between the upper bounds in (3.45) and (3.47), on the undetected error probability under the LR generalized decoding rule in (3.6). The comparison is provided for binary expurgated and regular (6,12) LDPC code ensembles of Gallager with block lengths of 504, 2004 and 10008 bits whose transmissions take place over binary-input AWGN channels with BPSK modulation.

event in this case corresponds to the possibility that the decoded list does not include the transmitted codeword.

Proposition 3.9 Consider the transmission of a block code $\mathcal{C}$ with M codewords of length n, and let $p(\mathbf{y}|\mathbf{x})$ designate the transition probability of the channel where $\mathbf{x} \in \mathcal{C}$ is the transmitted codeword and $\mathbf{y} \in \mathcal{Y}^n$ is the received vector. Consider the case where a fixed-size list decoder is used where the size of the list is denoted by L. Then, the conditional block error probability $P_{\mathrm{e}|m}$, given that the m-th message is transmitted satisfies

$$P_{\mathrm{e}|m} \leq \left(\sum_{\mathbf{y}} G_n^m(\mathbf{y}) p(\mathbf{y}|\mathbf{x}_m) \right)^{1-\rho} \left(\frac{1}{L} \sum_{m' \neq m} \sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}_m) G_N^m(\mathbf{y})^{1-\frac{1}{\rho}} \left(\frac{p(\mathbf{y}|\mathbf{x}_{m'})}{p(\mathbf{y}|\mathbf{x}_m)} \right)^{\frac{s}{\rho}} \right)^{\rho}. \tag{3.48}$$

where $0 \leq s \leq \rho \leq 1$ are real-valued parameters, and G_n^m is an arbitrary non-negative function over $\mathcal{Y}^n$ which possibly depends on the codeword $\mathbf{x}_m$, for $1 \leq m \leq M$.

Proof: See Appendix 3.I. ■

The following corollary is a particularization of Proposition 3.9 for the ensemble of fully-random block codes, with fixed block length and rate, whose transmission takes place over a memoryless channel:

Corollary 3.7 Consider the transmission of a block code $\mathcal{C}$ over a memoryless communication channel. Then, under the notation in Proposition 3.9, there exists a block code whose block error probability P_{e} under fixed-size list decoding satisfies

$$P_{\mathrm{e}} \leq e^{-nE_{\mathrm{r}}\left(R-\frac{1}{n}\ln L\right)} \tag{3.49}$$

where $R \triangleq \frac{\ln M}{n}$ is the code rate (in nats per channel use),

$$E_{\mathrm{r}}(R) \triangleq \max_{0\leq\rho\leq 1,\ q_X} \Big(E_0(\rho, q_X) - \rho R\Big) \tag{3.50}$$

$$E_0(\rho, q_X) \triangleq -\ln\left(\sum_{y\in\mathcal{Y}}\left(\sum_{x\in\mathcal{X}} q_X(x)p(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}\right)$$

and q_X is a probability distribution over the input alphabet $\mathcal{X}$.

Proof: Fix a probability distribution q_X over $\mathcal{X}$, and consider the ensemble of random block codes where each codeword is chosen independently according to $q_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{n} q_X(x_i)$. First, we apply the bound in (3.48) for a specific realization of a codebook, with $s = \frac{\rho}{1+\rho}$ and

$$G_n^m(\mathbf{y}) \triangleq \left(\sum_{\mathbf{x}} q_{\mathbf{X}}(\mathbf{x})\left(\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y}|\mathbf{x}_m)}\right)^{\frac{s}{\rho}}\right)^{\rho}.$$

The proof follows by a random coding argument, and by choosing the optimal probability distribution q_X. ■

Remark 3.15 (On comparison of the error exponent in Corollary 3.7 with previously known results) The upper bound in Corollary 3.7 is compared to three previously known results:

1. **The sphere-packing bound:** The sphere-packing lower bound in [97, eq. 1.6] provides an exponential lower bound on the error probability for fixed-size list-decoding of block codes. The bound in Corollary 3.7 and the sphere-packing bound exponentially coincide for all rates above the critical rate (where the maximization of the random coding error exponent is achieved for $0 \leq \rho \leq 1$).

2. **Asymptotic upper bound:** Consider the case where the size of the decoded list grows exponentially with the blocklength, and denote the exponential growth rate of the decoded list by l (i.e., $L = e^{nl}$ for some $l > 0$). The following asymptotic upper bound is provided in [24, p. 196, ex. 27] for the case at hand:

$$\limsup_{n\to\infty} \frac{1}{n} \ln P_e \leq -E_{\mathrm{r}}(R - l). \tag{3.51}$$

It is easily verified that the bound in Corollary 3.7 asymptotically coincides with the bound in (3.51).

3. **A variation on the Gallager bound:** The following exponential upper bound on the error probability is provided in [46, p. 538, ex. 5.20] for given block length and list size (the same assumptions and notation as in Corollary 3.7 are considered):

$$P_{\mathrm{e}} \leq e^{-nE_r(R,L)}$$

where

$$E_{\mathrm{r}}(R, L) \triangleq \max_{0\leq\rho\leq L,\ q_X} \Big(E_0(\rho, q_X) - \rho R \Big). \tag{3.52}$$

The error exponents in (3.49) and (3.52) differ in the following aspects:

(a) For a fixed list-size L, the error exponent in (3.49) depends on the block length n while the error exponent in (3.52) does not.

(b) The maximization of ρ in (3.49) is carried over the interval $[0, 1]$ while in (3.52) it is $[0, L]$.

(c) The bound in (3.49) includes an explicit rate reduction term, which depends on the list size.

(d) The derivation of the bound in (3.49) is based on a particularization of the DS2 bound in Proposition 3.9 for fully-random block codes. On the other hand, the derivation of the bound in (3.52) is based on a modification of the random coding bound [45] for the case at hand.

The two bounds in (3.49) and (3.52) are compared in Figure 3.13. Transmission of fully-random block codes over a BSC with a crossover of $p = 0.11$ are considered, where equiprobable $q_X(x) = \frac{1}{2}$, $x \in \mathcal{X}$, is assumed. The error exponent $E_{\mathrm{r}}(R, L)$ in (3.52) is plotted for a list size of $L = 16$ codewords. In addition, the exponent $E_{\mathrm{r}}(R - \frac{1}{n} \ln L)$ is provided for the same list size and block lengths of 128, 256 and 1024 bits. It is observed that for low rates the bound in (3.52) outperforms the bound in (3.49). For moderate rates, the bound in (3.49) outperforms the bound in (3.52). The gap between the plotted exponents is

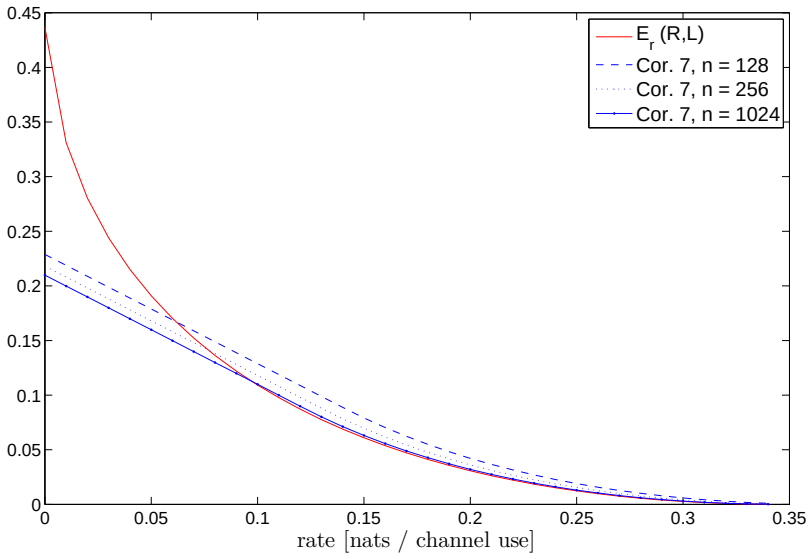


Figure 3.13: A comparison between the upper bounds in Corollary 3.7 and [46, p. 538, ex. 5.20]. Transmission of a fully-random binary block codes (with independent equiprobable selection of coded bits) over a BSC with a cross over probability of $p = 0.11$ is assumed. The exponent term $E_{\mathrm{r}}(R, L)$ in(3.52) is plotted for a list size of $L = 16$ codewords. The exponent $E_{\mathrm{r}}(R - \frac{1}{n} \ln L)$ in (3.49) is plotted for the same list-size and blocklengths of 128, 256 and 1024 bits.

negligible as the block length increases (even for a moderate block length of 1024 bits).

The following bound is provided for the case of binary linear block codes whose transmission takes place over an MBIOS channel:

Corollary 3.8 Consider an (n, k) binary linear block code $\mathcal{C}$ whose transmission takes place over an MBIOS channel. Then, the block error probability P_{e} under fixed-size list-decoding, satisfies

$$P_{\mathrm{e}} \leq e^{-nE_{\mathrm{r}}\left(R + \frac{1}{n} \ln\left(\frac{\alpha(\mathcal{C})}{L}\right)\right)} \tag{3.53}$$

where

$$E_{\mathrm{r}}(R) \triangleq \max_{0 \leq \rho \leq 1} \big(E_0(\rho) - \rho R\big)$$

and R is the code rate (in nats per channel use), L is the list size, and $E_0(\rho)$ and $\alpha(\mathcal{C})$ are defined in (3.21) and (3.22), respectively.

Proof: According to Proposition 3.4, it is necessary to analyze only the conditional error event assuming that the all-zero codeword is transmitted. Setting

$G_n^0(\mathbf{y}) = \prod_{i=1}^n g(y_i)$ in (3.48), it follows that

$$P_{\mathrm{e}} \leq \left(\sum_{y\in\mathcal{Y}} g(y)p(y|0)\right)^{n(1-\rho)} \left(\frac{1}{L}\sum_{i=1}^{n} |\mathcal{C}_i| \left(\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}} p(y|0)\right)^{n-i} \left(\sum_{y\in\mathcal{Y}} g(y)^{1-\frac{1}{\rho}} p(y|1)^{\lambda} p(y|0)^{1-\lambda}\right)^{i}\right)^{\rho} \quad (3.54)$$

where $|\mathcal{C}_i|$ denotes the number of codewords whose Hamming distance is i, $1 \leq i \leq n$. The proof follows from (3.54) by setting $\lambda = \frac{1}{1+\rho}$ where g is as defined in (3.23) (see similar derivation in [94, Section 4.4.1]). ■

Remark 3.16 For the particular case of fully-random linear block codes, the bound in (3.53) coincides with the bound in Corollary 3.7 for fully-random block codes.

Remark 3.17 The bound in Corollary 3.8 resembles to the SFB [100], and therefore may be considered as a generalization of the SFB for the case at hand.

Remark 3.18 The bound in (3.54) can be generalized to non-binary linear block codes using a similar derivation as in Chapter 2. Note, however, that in Chapter 2, non-binary codes are studied under ML decoding and not list-decoding. Nevertheless, the similarity of the bound in (3.48) to the upper bounds derived in Chapter 2 allows to use the same arguments for the case at hand (see Appendix 3.I).

Corollary 3.9 Under the assumptions and notation in Corollary 3.3, the block error probability probability P_{e} under fixed-size list-decoding where L denotes the size of the list, satisfies

$$P_{\mathrm{e}} \leq A(\rho)^{n(1-\rho)} \left(\frac{1}{L}\sum_{1\leq l\leq n} P(l)\binom{n}{l} B(\rho)^{n-l} C(\rho)^{l}\right)^{\rho} \quad (3.55)$$

where $A(\rho)$, $B(\rho)$, and $C(\rho)$ are defined in (3.29)–(3.31).

Proof: Setting $s = \frac{\rho}{1+\rho}$ and $G_n^m(\mathbf{y}) = \prod_{i=1}^n g(y_i)$ where g is defined in (3.23), the proof follows from Proposition 3.9 in the same way as the proof of Theorem 2.3. ■

Remark 3.19 In the derivation of the bound in (3.53), a sum is upper bounded by a product of the maximal summand with the number of summands. This operation is avoided in the derivation of the bound in (3.55). Hence, the bound in Corollary 3.9 is tighter than the one in Corollary 3.8.

Remark 3.20 For the particular case of binary linear block codes, the symmetry condition in (3.25) is not mandatory and the bound in Corollary 3.9 follows by replacing the term $P(l)\binom{n}{l}$ with the distance spectrum of the considered code (ensemble).

Example 3.11 (Error performance of an expurgated ensemble of binary and regular LDPC codes under fixed-size list decoding) Consider the expurgation of Gallager's ensemble of binary and regular (6,12) LDPC codes with a block length of 2004 bits (see Example 3.3). Upper bounds on the block error probability under fixed-size list-decoding are shown in Figures 3.14(a) and 3.14(b), assuming that the transmission takes place over a BSC and a binary-input AWGN channel, respectively. The upper bound in Corollary 3.9 is evaluated for list sizes of $L = 1$, 16, and 128 codewords. Note that the upper bound for $L = 1$ corresponds to ML decoding. The bounds on the error probability show some marginal improvement by increasing the considered list size from $L = 1$ to 128.

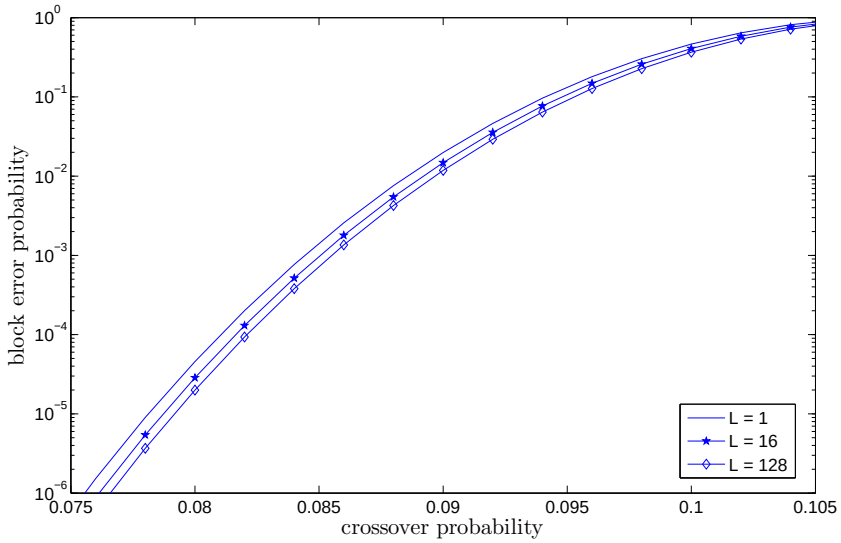
Example 3.12 (Error performance of an expurgated ensemble of non-binary and regular LDPC codes under fixed-size list decoding) Consider the expurgation of Gallager's ensemble of regular (8,16) LDPC codes with octal alphabet and a block length of 1008 symbols (see Example 3.5). Upper bounds on the block error probability under fixed-size list decoding are shown in Figures 3.15(a) and 3.15(b), assuming that the transmission takes place over an 8-ary discrete memoryless symmetric channel and an AWGN channel with 8-PSK modulation, respectively. The bound in Corollary 3.9 is evaluated for list sizes of $L = 1$, 16, and 128 codewords. similarly to the case of binary code ensembles, only marginal improvement in the error performance is observed by increasing the value of L from 1 to 128.

3.6 Summary and Conclusions

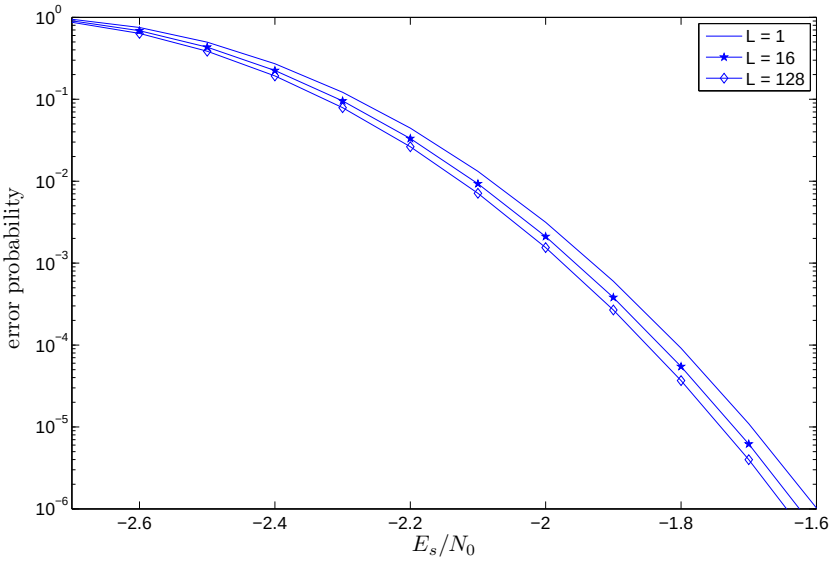
This chapter considers performance bounds for several generalized decoding rules over memoryless symmetric channels. Three types of generalized decoding rules are considered:

1. The optimal generalized decoding rule in [41] with erasures and variable list sizes.

2. The suboptimal likelihood-ratio (LR) decoding rule with erasures (see [9] and [41]).

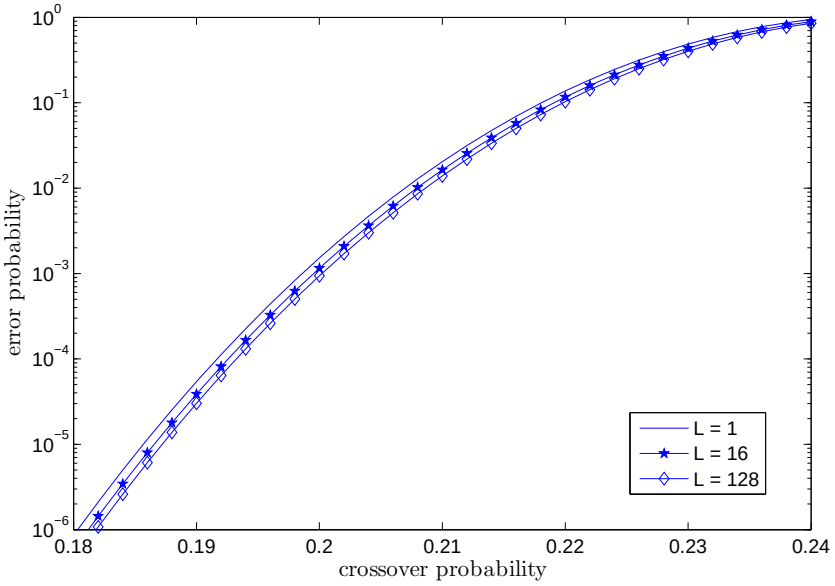


(a) Transmission over a BSC

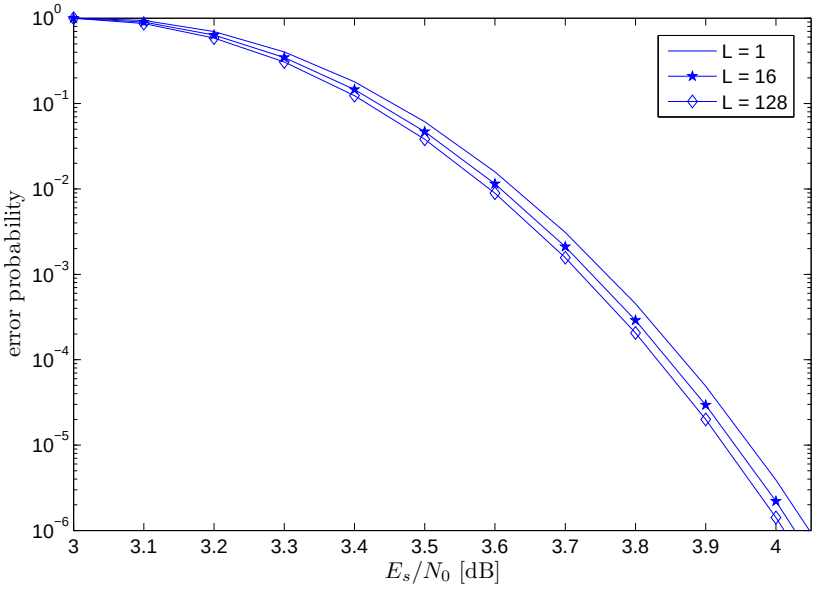


(b) Transmission over a binary-input AWGN channel

Figure 3.14: Upper bounds on the error probability for an expurgation of Gallager's ensemble of binary and regular (6,12) LDPC codes with a block length of 2004 bits (see Example 3.3). A list decoder is assumed where the size of the list is set to L. The upper bound in Corollary 3.9 is provided for some values of L. The bounds are shown in plots (a) and (b), respectively, for the case where the transmission takes place over a BSC and a binary-input AWGN channel.



(a) Transmission over an 8-ary discrete memoryless symmetric channel



(b) Transmission over an AWGN channel with 8-PSK modulation

Figure 3.15: Upper bounds on the error probability for an expurgation of Gallager's ensemble of regular (8,16) LDPC codes with octal alphabet and a block length of 1008 symbols (see Example 3.5). A list decoder is considered where the size of the list is set to L. The upper bound in Corollary 3.9 is provided in plots (a) and (b) for several values of L, assuming that the transmission takes place over an 8-ary discrete memoryless symmetric channel and an AWGN channel with 8-PSK modulation, respectively.

3. A fixed-size list decoding rule (see [36] and [117]) where the decoder outputs a list which includes the L most probable codewords (where the value of L is set a-priori).

The independence of the error performance on the transmitted codeword is proved in Propositions 3.2-3.4 for the considered decoding rules.

Upper bounds on the decoding error probability are provided. These bounds are suitable for the analysis of structured and random codes (or code ensembles) over memoryless symmetric channels. Both binary and non-binary code ensembles are studied in this chapter under generalized decoding rules. When binary codes are considered, the bounds are based on the distance spectra of the codes, and when non-binary ensembles are studied, the complete composition spectra are required under the symmetry assumption in (3.25). For the case of LR decoding of binary linear block codes, a derivation of a closed-form expression is provided via a similar derivation to [26] which applies to ML decoding.

Several particular cases of the provided bounds are studied. The random coding error exponents in [41] are reproduced. In addition, error exponents under the suboptimal LR decoding rule with erasures are also derived. These error exponents are derived by applying the new bounds to fully random block codes. Next, a derivation of the error exponents of fully random linear block codes under optimal and suboptimal (LR) generalized decoding is provided. The resulting error exponents under the suboptimal LR decoding rule are compared with a recent improvement in [9], where the ensemble of binary fully random linear block codes over binary symmetric channels (BSC) is studied. This comparison shows good match with the provided error exponents with the results in [9]. In addition, it is shown that the error exponents for the fully random linear block codes under the suboptimal LR decoding rule, coincide for low rates with the corresponding error exponents under the optimal decoding rule. This is similar to an observation in [54], where the ensemble of fully random block codes is considered. A lower bound on the error exponent under fixed-size list-decoding is also studied as an application. This bound is compared to the sphere-packing lower bound on the error probability [97], and two additional upper bounds on the error probability, provided in [24] and [46].

Applications of the bounds for the performance analysis of structured code ensembles are further exemplified for some expurgated ensembles of (binary and non-binary) regular low-density parity-check (LDPC) codes. The error performance under some generalized decoding rules for these LDPC code ensembles is studied assuming that the transmission takes place over memoryless symmetric channels. The application of the provided bounds for the study of hybrid automatic-repeat request (ARQ)

schemes is also demonstrated. The possibility of further investigating and optimizing the trade-offs between undetected error and erasures is suggested for further study in the context of linear block codes, based on the derived bounds.

Appendices

3.A Proof of Proposition 3.2

The following proof holds for memoryless symmetric channels with discrete-output alphabets, and the generalization to continuous-output alphabets is direct.

Assuming that all the codewords are sent with equal probability, the decision regions in (3.4) satisfy

$$\begin{aligned}\Lambda_m &\overset{(a)}{=} \left\{\mathbf{y}:\ \frac{p(\mathbf{y}|\mathbf{x}_m)}{\sum_{m'\neq m} p(\mathbf{y}|\mathbf{x}_{m'})} \geq e^{nT}\right\} \\ &\overset{(b)}{=} \left\{\mathbf{y}:\ \frac{\prod_{i=1}^n p(y_i|x_{m,i})}{\sum_{m'\neq m}\prod_{i=1}^n p(y_i|x_{m',i})} \geq e^{nT}\right\} \\ &\overset{(c)}{=} \left\{\mathbf{y}:\ \frac{\prod_{i=1}^n p(\mathcal{T}(y_i,-x_{m,i})|0)}{\sum_{m'\neq m}\prod_{i=1}^n p(\mathcal{T}(y_i,-x_{m',i})|0)} \geq e^{nT}\right\} \end{aligned} \tag{3.A.1}$$

where (a) follows from (3.4) and the equal a-priori message probability assumption, (b) holds since the channel is memoryless, and (c) follows from the symmetry of the channel (see (2.1)). Let $\mathbf{z} = (z_1, \ldots, z_n)$ be defined as

$$z_i \triangleq \mathcal{T}(y_i, -x_{m,i}), \quad 1 \leq i \leq n \tag{3.A.2}$$

where m is the index of the transmitted codeword. From Lemma 2.1, it follows that $\mathbf{y} \in \Lambda_m$ if and only if $\mathbf{z} \in \tilde{\Lambda}_m$ where

$$\tilde{\Lambda}_m \triangleq \left\{\mathbf{z}\in\mathcal{Y}^n:\ \frac{\prod_{i=1}^n p(z_i|0)}{\sum_{m'\neq m}\prod_{i=1}^n p(\mathcal{T}(z_i, x_{m,i}-x_{m',i})|0)} \geq e^{nT}\right\}, \quad 1\leq m\leq q^k.$$

Using the linearity of the code, it follows that

$$\tilde{\Lambda}_m = \left\{\mathbf{z}\in\mathcal{Y}^n:\ \frac{\prod_{i=1}^n p(z_i|0)}{\sum_{l\neq 0}\prod_{i=1}^n p(\mathcal{T}(z_i, x_{l,i})|0)} \geq e^{nT}\right\}.$$

Since the set $\tilde{\Lambda}_m$ is independent of the index m, then

$$\tilde{\Lambda}_m = \tilde{\Lambda}_1 \text{ for all } 1\leq m\leq q^k. \tag{3.A.3}$$

As a result, the conditional block error probability of the m-th message in (3.1) satisfies

$$\begin{aligned} P_{\mathrm{e}|m} &= \sum_{\mathbf{z}\in\tilde{\Lambda}_m^{\mathrm{c}}} p(\mathbf{z}|\mathbf{0}) \\ &\overset{(a)}{=} \sum_{\mathbf{z}\in\tilde{\Lambda}_1^{\mathrm{c}}} p(\mathbf{z}|\mathbf{0}) \end{aligned}$$

where (a) follows from (3.A.3). This concludes the proof of the message independence property for the block error event.

We continue in proving the message independence property for the undetected error event (or the expected number of incorrect codewords when list decoding is considered). Assuming a memoryless symmetric channel, it follows from (2.1) and (3.3) that

$$\begin{aligned} P_{\mathrm{ue}|m} &= \sum_{m'\neq m} \sum_{\mathbf{y}\in\Lambda_{m'}} p(\mathbf{y}|\mathbf{x}_m) \\ &= \sum_{m'\neq m} \sum_{\mathbf{y}\in\Lambda_{m'}} \prod_{i=1}^{n} p\big(\mathcal{T}(y_i, -x_{m,i})|0\big) \end{aligned} \tag{3.A.4}$$

where from (3.A.1)

$$\Lambda_{m'} = \left\{ \mathbf{y} : \ \frac{\prod_{i=1}^{n} p(\mathcal{T}(y_i, -x_{m',i})|0)}{\sum_{m''\neq m'} \prod_{i=1}^{n} p(\mathcal{T}(y_i, -x_{m'',i})|0)} \geq e^{nT} \right\}.$$

Let $\mathbf{z}$ be a vector defined as in (3.A.2), then from Lemma 2.1

$$p\big(\mathcal{T}(y_i, -x_{m',i})|0\big) = p\big(\mathcal{T}(z_i, x_{m,i} - x_{m',i}|0\big), \quad i = 1, \ldots, n.$$

Hence, given that $\mathbf{x}_m$ is the transmitted codeword, then $\mathbf{y} \in \Lambda_{m'}$ for some $m' \neq m$ if and only if $\mathbf{z} \in \Gamma_{m,m'}$ where

$$\Gamma_{m,m'} \triangleq \left\{ \mathbf{z} \in \mathcal{Y}^n : \ \frac{\prod_{i=1}^{n} p(\mathcal{T}(z_i, x_{m,i} - x_{m',i})|0)}{\sum_{m''\neq m'} \prod_{i=1}^{n} p(\mathcal{T}(z_i, x_{m,i} - x_{m'',i})|0)} \geq e^{nT} \right\}. \tag{3.A.5}$$

From (3.A.2), the conditional undetected error probability in (3.A.4) is rewritten in the form

$$P_{\mathrm{ue}|m} = \sum_{m'\neq m} \sum_{\mathbf{z}\in\Gamma_{m,m'}} p(\mathbf{z}|\mathbf{0}). \tag{3.A.6}$$

Using the linearity of the code, then $x_{m,i} - x_{m'',i} = \big(x_{m,i} - x_{m',i}\big) + \big(x_{m',i} - x_{m'',i}\big) = x_{l_1,i} + x_{l_2,i}$ for some indices l_1 and l_2 which correspond to non-zero codewords. Let

$\mathbf{x} \triangleq \mathbf{x}_{l_1}$ and $\tilde{\mathbf{x}} = \mathbf{x}_{l_2}$, then the conditional undetected error probability in (3.A.6) is expressed equivalently in the form

$$P_{\mathrm{ue}|m} = \sum_{\substack{\mathbf{x}\in\mathcal{C}\\ \mathbf{x}\neq\mathbf{0}}} \sum_{\mathbf{z}\in\Gamma(\mathbf{x})} p(\mathbf{z}|\mathbf{0})$$

where, based on (3.A.5),

$$\Gamma(\mathbf{x}) \triangleq \left\{ \mathbf{z}\in\mathcal{Y}^n : \ \frac{\prod_{i=1}^n p(\mathcal{T}(z_i,x_i)|0)}{\sum_{\substack{\tilde{\mathbf{x}}\in\mathcal{C}\\ \tilde{\mathbf{x}}\neq\mathbf{0}}} \prod_{i=1}^n p(\mathcal{T}(z_i,x_i+\tilde{x}_i)|0)} \geq e^{nT} \right\}.$$

This proves the independence property for the undetected error event, and it concludes the proof of Proposition 3.2.

3.B Proof of Proposition 3.3

Similarly to Appendix 3.A, also the following proof considers memoryless symmetric channels with discrete-output alphabets, where the generalization to continuous output alphabets is direct. Let p be the transition probability function of the considered channel, $\mathcal{C}$ be an (n,k) linear block code over an alphabet whose cardinality is q, and $\mathcal{T}$ be a mapping as specified in Definition 2.1. It is assumed that all the codewords of $\mathcal{C}$ are sent with equal probability. For an arbitrary set $\Lambda \subseteq \mathcal{Y}^n$ and a codeword $\mathbf{x}_m \in \mathcal{C}$, let

$$\mathcal{Z}_m(\Lambda) \triangleq \Big\{ \mathbf{z}\in\mathcal{Y}^n : \ \big(\mathcal{T}(z_1,x_{m,1}), \mathcal{T}(z_2,x_{m,2}), \ldots, \mathcal{T}(z_n,x_{m,n})\big) \in \Lambda \Big\}. \tag{3.B.7}$$

In addition, we use the notation $\Lambda^{\mathrm{LR}}(\mathbf{x}_m)$ for the decision region Λ_m^{LR} in (3.6) of the codeword $\mathbf{x}_m$. Note that for the concerned decoding rule with $T > 0$, the decision regions are disjoint. The following technical lemma is introduced:

Lemma 3.B.1 Let $\mathcal{Z}_m$ be the mapping defined in (3.B.7), and Λ_m^{LR} be the decision region in (3.6). Then,

$$\mathcal{Z}_m\left(\Lambda_{m'}^{\mathrm{LR}}\right) = \Lambda^{\mathrm{LR}}(\mathbf{x}_{m'} - \mathbf{x}_m), \quad \forall\, m,m' \in \{1,\ldots,q^k\}. \tag{3.B.8}$$

Proof: Let us choose $\mathbf{z} \in \mathcal{Z}_m\left(\Lambda_{m'}^{\mathrm{LR}}\right)$, and let $\mathbf{y} = (y_1,\ldots,y_n)$ be defined via the equality

$$y_i = \mathcal{T}(z_i, x_{m,i}), \quad i = 1,\ldots,n. \tag{3.B.9}$$

From (3.6) and (3.B.7)

$$\frac{p(\mathbf{y}|\mathbf{x}_{m'})}{p(\mathbf{y}|\mathbf{x}_{m_2'})} \geq e^{nT}$$

where $\mathbf{x}_{m'}$ and $\mathbf{x}_{m_2'}$ are the most probable codewords, in a descending order, for $\mathbf{y}$ as a received vector. Using the symmetry of the channel, it follows from (2.1) that

$$p(\mathbf{y}|\mathbf{x}_{m'}) = p(\mathbf{z}|\mathbf{x}_{m'} - \mathbf{x}_m).$$

As a result, $\mathbf{x}_{m'} - \mathbf{x}_m$ is the most probable codeword if $\mathbf{z}$ is the received vector (otherwise, if there exists a codeword $\mathbf{x} \neq \mathbf{x}_{m'} - \mathbf{x}_m$ which is more probable, then there exists a more probable codeword for $\mathbf{y}$ which is different from $\mathbf{x}_{m'}$). The same argument shows that $\mathbf{x}_{m_2'} - \mathbf{x}_m$ is the second most probable codeword for $\mathbf{z}$, and

$$\frac{p(\mathbf{z}|\mathbf{x}_{m'} - \mathbf{x}_m)}{p(\mathbf{z}|\mathbf{x}_{m_2'} - \mathbf{x}_m)} \geq e^{nT}.$$

This verifies that $\mathbf{z} \in \Lambda^{\mathrm{LR}}(\mathbf{x}_{m'} - \mathbf{x}_m)$ which shows that $\mathcal{Z}_m\left(\Lambda_{m'}^{\mathrm{LR}}\right) \subseteq \Lambda^{\mathrm{LR}}(\mathbf{x}_{m'} - \mathbf{x}_m)$. To show the opposite inclusion, which then yields that these two sets are equal, let $\mathbf{z} \in \Lambda^{\mathrm{LR}}(\mathbf{x}_{m'} - \mathbf{x}_m)$. This implies that the codeword $\mathbf{x}_{m'} - \mathbf{x}_m$ is the most probable codeword if $\mathbf{z}$ is the received vector, and

$$\frac{p(\mathbf{z}|\mathbf{x}_{m'} - \mathbf{x}_m)}{p(\mathbf{z}|\mathbf{x}_{m_2''})} \geq e^{nT}$$

where $\mathbf{x}_{m_2''}$ is the second most probable codeword for $\mathbf{z}$. Again, using the symmetry of the channel, for a vector $\mathbf{y}$ as in (3.B.9), it follows that $\mathbf{x}_{m'}$ is the most probable codeword for $\mathbf{y}$, $\mathbf{x}_{m_2''} + \mathbf{x}_m$ is the second most probable codeword for $\mathbf{y}$, and

$$\frac{p(\mathbf{y}|\mathbf{x}_{m'})}{p(\mathbf{y}|\mathbf{x}_{m_2''} + \mathbf{x}_m)} \geq e^{nT}.$$

As a result, $\mathbf{z} \in \mathcal{Z}_m\left(\Lambda_{m'}^{\mathrm{LR}}\right)$, which yields that $\Lambda^{\mathrm{LR}}(\mathbf{x}_{m'} - \mathbf{x}_m) \subseteq \mathcal{Z}_m\left(\Lambda_{m'}^{\mathrm{LR}}\right)$. This concludes the proof of (3.B.8). ■

From (3.B.9), the conditional block error probability satisfies

$$\begin{aligned} P_{\mathrm{e}|m} &= \sum_{\mathbf{y} \notin \Lambda_m^{\mathrm{LR}}} p(\mathbf{y}|\mathbf{x}_m) \\ &\overset{\text{(a)}}{=} \sum_{\mathbf{z} \notin \mathcal{Z}_m(\Lambda_m^{\mathrm{LR}})} p(\mathbf{z}|\mathbf{0}) \\ &\overset{\text{(b)}}{=} \sum_{\mathbf{z} \notin \Lambda^{\mathrm{LR}}(\mathbf{0})} p(\mathbf{z}|\mathbf{0}) \end{aligned}$$

where (a) follows from (2.1) and (3.B.9), and (b) follows from (3.B.8). This proves the message independence property for the conditional block error probability. Using the

same arguments, the message independence property is established for the conditional undetected error probability:

$$\begin{aligned} P_{\mathrm{ue}|m} &= \sum_{m' \neq m} \sum_{\mathbf{y} \in \Lambda_{m'}^{\mathrm{LR}}} p(\mathbf{y}|\mathbf{x}_m) \\ &= \sum_{m' \neq m} \sum_{\mathbf{z} \in \mathcal{Z}_m\left(\Lambda_{m'}^{\mathrm{LR}}\right)} p(\mathbf{z}|\mathbf{0}) \\ &= \sum_{m' \neq m} \sum_{\mathbf{z} \in \Lambda^{\mathrm{LR}}(\mathbf{x}_{m'} - \mathbf{x}_m)} p(\mathbf{z}|\mathbf{0}) \\ &= \sum_{\substack{\mathbf{x} \in \mathcal{C} \\ \mathbf{x} \neq \mathbf{0}}} \sum_{\mathbf{z} \in \Lambda^{\mathrm{LR}}(\mathbf{x})} p(\mathbf{z}|\mathbf{0}) \end{aligned}$$

where the second equality follows from (3.B.9) and since the mapping $\mathcal{T}$ is bijective, the third equality follows from (3.B.8), and the last equality follows from the linearity of the code.

3.C Proof of Proposition 3.4

Considering ties as error events[1], the conditional block error probability for a list of size L satisfies

$$P_{\mathrm{e}|m} = \sum_{\mathbf{y} \in \Lambda_m^L} p(\mathbf{y}|\mathbf{x}_m) \tag{3.C.10}$$

where

$$\Lambda_m^L \triangleq \left\{ \mathbf{y} \in \mathcal{Y}^n : \ \exists \{m_i\}_{i=1}^L \text{ s.t. } m_i \neq m, \ p(\mathbf{y}|\mathbf{x}_{m_i}) \geq p(\mathbf{y}|\mathbf{x}_m) \ \forall \ 1 \leq i \leq L \right\} \tag{3.C.11}$$

is the complementary of the decision region of $\mathbf{x}_m \in \mathcal{C}$ under list decoding of fixed-size L (here $\{m_i\}_{i=1}^L$ is a sequence of distinct integers), i.e., if $\mathbf{y} \in \Lambda_m^L$ then the codeword $\mathbf{x}_m$ is not included in the list for a received vector $\mathbf{y}$. Using the change of variables in (3.B.9), it follows from (3.C.10) that for linear block codes whose transmission takes place over memoryless symmetric channels

$$P_{\mathrm{e}|m} = \sum_{\mathbf{z} \in \mathcal{Z}_m(\Lambda_m^L)} p(\mathbf{z}|\mathbf{0})$$

where $\mathcal{Z}_m\left(\Lambda_m^L\right)$ is as defined in (3.B.7). The following lemma concludes the proof of Proposition 3.4:

[1]Such a pessimistic assumption is reasonable, see also a similar assumption in [111, p. 59].

Lemma 3.C.2 Let $\mathcal{Z}_m$ be a mapping defined in (3.B.7), and Λ_m^L be the decoding region of $\mathbf{x}_m \in \mathcal{C}$ under list decoding with a fixed size L. Then,

$$\mathcal{Z}_m\left(\Lambda_m^L\right) = \Lambda_1^L$$

for all $1 \leq m \leq q^k$, where Λ_1^L is the complementary of the decision region of the all-zero codeword $\mathbf{x}_1 = \mathbf{0}$ under list decoding of size L.

Proof: Let us choose $\mathbf{z} \in \mathcal{Z}\left(\Lambda_m^L\right)$. From (3.B.7), there exists $\mathbf{y} \in \Lambda_m^L$ where

$$y_i = \mathcal{T}(z_i, x_{m,i}), \quad i = 1, \ldots, n \tag{3.C.12}$$

and $\mathcal{T}$ is a specified in Definition 2.1. From (3.C.11), there exists a list of L distinct codewords, $\{\mathbf{x}_{m_i}\}_{i=1}^L$, for which

$$p(\mathbf{y}|\mathbf{x}_{m_i}) > p(\mathbf{y}|\mathbf{x}_m), \quad i = 1, \ldots, L. \tag{3.C.13}$$

Using the symmetry of the channel, it follows that

$$p(\mathbf{z}|\mathbf{x}_{m_i} - \mathbf{x}_m) \geq p(\mathbf{z}|\mathbf{0}). \tag{3.C.14}$$

This assures that $\mathbf{z} \in \Lambda_1^L$, which shows that $\mathcal{Z}_m\left(\Lambda_m^L\right) \subseteq \Lambda_1^L$.

Next, in order to show the opposite inclusion, let $\mathbf{z} \in \Lambda_1^L$. Then, there exists a list of L non-zero codewords $\{\mathbf{x}_{m_i}\}_{i=1}^L$, $m_i \neq 1$, satisfying

$$p(\mathbf{z}|\mathbf{x}_{m_i}) \geq p(\mathbf{z}|\mathbf{0})$$

and therefore from the symmetry of the mapping $\mathcal{T}$ and the equality in (3.C.12), we get

$$p(\mathbf{y}|\mathbf{x}_{m_i} + \mathbf{x}_m) \geq p(\mathbf{y}|\mathbf{x}_m)$$

It assures that $\mathbf{z} \in \mathcal{Z}_m\left(\Lambda_m^L\right)$ which implies that $\Lambda_1^L \subseteq \mathcal{Z}_m\left(\Lambda_m^L\right)$. This two inclusions complete the proof of the lemma. ■

3.D Proof of Proposition 3.5

Let Λ_m be the generalized decision region as defined in (3.4). For $\mathbf{y} \notin \Lambda_m$, it follows that

$$1 = e^{nT}e^{-nT} \leq e^{nT}\left(\sum_{m' \neq m} \frac{p(\mathbf{y}|\mathbf{x}_{m'})}{p(\mathbf{y}|\mathbf{x}_m)}\right). \tag{3.D.15}$$

Let s and ρ satisfy $0 \leq s \leq \rho \leq 1$, and recall the following inequality (see [111, p.197]):

$$\sum_i a_i \leq \left(\sum_i a_i^\lambda \right)^{\frac{1}{\lambda}} \tag{3.D.16}$$

which holds if $a_i \geq 0$ and $0 < \lambda \leq 1$. Setting

$$a_i = \frac{p(\mathbf{y}|\mathbf{x}_i)}{p(\mathbf{y}|\mathbf{x}_m)}, \quad \lambda = \frac{s}{\rho}$$

it follows from (3.1), (3.D.15) and (3.D.16) that the conditional error probability of the m-th message satisfies

$$\begin{aligned} P_{\mathrm{e}|m} &\leq e^{nTs} \sum_{\mathbf{y} \in \Lambda_m^c} p(\mathbf{y}|\mathbf{x}_m) \left(\sum_{m' \neq m} \frac{p(\mathbf{y}|\mathbf{x}_{m'})}{p(\mathbf{y}|\mathbf{x}_m)} \right)^s \\ &\leq e^{nTs} \sum_{\mathbf{y} \in \Lambda_m^c} p(\mathbf{y}|\mathbf{x}_m) \left(\sum_{m' \neq m} \left(\frac{p(\mathbf{y}|\mathbf{x}_{m'})}{p(\mathbf{y}|\mathbf{x}_m)} \right)^{\frac{s}{\rho}} \right)^\rho . \end{aligned} \tag{3.D.17}$$

Let $\psi_n^m(\mathbf{y})$ designate an arbitrary probability tilting measure (which may depend on the transmitted codeword), then it follows that

$$\begin{aligned} P_{\mathrm{e}|m} &\leq e^{nTs} \sum_{\mathbf{y}} \psi_n^m(\mathbf{y}) \psi_n^m(\mathbf{y})^{-1} p(\mathbf{y}|\mathbf{x}_m) \left(\sum_{m' \neq m} \left(\frac{p(\mathbf{y}|\mathbf{x}_{m'})}{p(\mathbf{y}|\mathbf{x}_m)} \right)^{\frac{s}{\rho}} \right)^\rho \\ &\leq e^{nTs} \sum_{\mathbf{y}} \psi_n^m(\mathbf{y}) \left(\psi_n^m(\mathbf{y})^{-\frac{1}{\rho}} p(\mathbf{y}|\mathbf{x}_m)^{\frac{1}{\rho}} \sum_{m' \neq m} \left(\frac{p(\mathbf{y}|\mathbf{x}_{m'})}{p(\mathbf{y}|\mathbf{x}_m)} \right)^{\frac{s}{\rho}} \right)^\rho . \end{aligned}$$

Next, invoking Jensen's inequality gives

$$P_{\mathrm{e}|m} \leq e^{nTs} \left(\sum_{\mathbf{y}} \psi_n^m(\mathbf{y})^{1-\frac{1}{\rho}} p(\mathbf{y}|\mathbf{x}_m)^{\frac{1}{\rho}} \sum_{m' \neq m} \left(\frac{p(\mathbf{y}|\mathbf{x}_{m'})}{p(\mathbf{y}|\mathbf{x}_m)} \right)^{\frac{s}{\rho}} \right)^\rho .$$

This concludes the proof of (3.8) by setting

$$\psi_n^m(\mathbf{y}) = \frac{G_n^m(\mathbf{y}) p(\mathbf{y}|\mathbf{x}_m)}{\sum_{\mathbf{y}} G_n^m(\mathbf{y}) p(\mathbf{y}|\mathbf{x}_m)} \tag{3.D.18}$$

where $G_n^m(\mathbf{y})$ is an arbitrary non-negative function.

An undetected error event occurs if the received vector is included in the decision region of a codeword which differs from the transmitted codeword. Consequently, the average undetected error event satisfies

$$P_{\mathrm{ue}} = \frac{1}{M} \sum_{m=1}^{M} \sum_{\mathbf{y} \in \Lambda_m} \sum_{m' \neq m} p(\mathbf{y}|\mathbf{x}_{m'}). \tag{3.D.19}$$

Note that in the case where list decoding is considered (i.e., the decision regions are not disjoint), the LHS of (3.D.19) is no longer a probability. However, for the latter case this expression equals the expected number of incorrect codewords in the decoded list. It follows from (3.D.19) that for $0 \leq s \leq 1$, the undetected error probability satisfies

$$\begin{aligned} P_{\text{ue}} &= \frac{1}{M} \sum_{m=1}^{M} \sum_{\mathbf{y} \in \Lambda_m} p(\mathbf{y}|\mathbf{x}_m) \left(\frac{\sum_{m' \neq m} p(\mathbf{y}|\mathbf{x}_{m'})}{p(\mathbf{y}|\mathbf{x}_m)} \right)^s \left(\frac{\sum_{m' \neq m} p(\mathbf{y}|\mathbf{x}_{m'})}{p(\mathbf{y}|\mathbf{x}_m)} \right)^{1-s} \\ &\leq e^{nT(s-1)} \frac{1}{M} \sum_{m=1}^{M} \sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}_m) \left(\sum_{m' \neq m} \frac{p(\mathbf{y}|\mathbf{x}_{m'})}{p(\mathbf{y}|\mathbf{x}_m)} \right)^s \end{aligned} \tag{3.D.20}$$

where the last inequality holds since for $\mathbf{y} \in \Lambda_m$ and $0 \leq s \leq 1$

$$\left(\frac{p(\mathbf{y}|\mathbf{x}_m)}{\sum_{m' \neq m} p(\mathbf{y}|\mathbf{x}_{m'})} \right)^{1-s} \geq e^{nT(1-s)}.$$

The rest of the proof follows in a similar way to the derivation of (3.8) when comparing the bound in (3.D.17) with (3.D.20).

3.E Proof of Corollary 3.1

Consider the ensemble of fully random block codes of length n symbols where the $M = e^{nR}$ codewords of a codebook are chosen independently at random according to the probability distribution $q_{\mathbf{X}}$ on $\mathcal{X}^n$.

Let $D_{\{\mathbf{x}_i\}_{i=1}^M}(m, G_n^m, s, \rho)$ denote the functional $D_{\text{B}}(m, G_n^m, s, \rho)$ in (3.10) where the dependence on a specific codebook $\{\mathbf{x}_i\}_{i=1}^M$ is expressed explicitly. Given a fixed codeword $\mathbf{x}_m$ for the m-th message, the expectation over the other $M-1$ codewords on the right-hand side of (3.8) gives that for $0 \leq s \leq \rho \leq 1$

$$\begin{aligned} &\sum_{\{\mathbf{x}_i\}_{i=1}^M \setminus \{\mathbf{x}_m\}} \left(\prod_{i \neq m} q_{\mathbf{X}}(\mathbf{x}_i) \right) D_{\{\mathbf{x}_i\}_{i=1}^M}(m, G_n^m, s, \rho) \\ &\overset{(a)}{\leq} \left(\sum_{\mathbf{y}} G_n^m(\mathbf{y}) p(\mathbf{y}|\mathbf{x}_m) \right)^{1-\rho} \\ &\quad \left(\sum_{m' \neq m} \sum_{\mathbf{x}_{m'}} q_{\mathbf{X}}(\mathbf{x}_{m'}) \sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}_m) G_N^m(\mathbf{y})^{1-\frac{1}{\rho}} \left(\frac{p(\mathbf{y}|\mathbf{x}_{m'})}{p(\mathbf{y}|\mathbf{x}_m)} \right)^{\frac{s}{\rho}} \right)^{\rho} \end{aligned}$$

$$= (M-1)^{\rho} \left(\sum_{\mathbf{y}} G_n^m(\mathbf{y}) p(\mathbf{y}|\mathbf{x}_m) \right)^{1-\rho}$$
$$\left(\sum_{\mathbf{x}'} q_{\mathbf{X}}(\mathbf{x}') \sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}_m) G_N^m(\mathbf{y})^{1-\frac{1}{\rho}} \left(\frac{p(\mathbf{y}|\mathbf{x}')}{p(\mathbf{y}|\mathbf{x}_m)} \right)^{\frac{s}{\rho}} \right)^{\rho} \tag{3.E.21}$$

where (a) follows from (3.10) and by invoking Jensen's inequality. Next, by substituting the non-negative function

$$G_n^m(\mathbf{y}) \triangleq \left(\sum_{\mathbf{x}} q_{\mathbf{X}}(\mathbf{x}) \left(\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y}|\mathbf{x}_m)} \right)^{\frac{s}{\rho}} \right)^{\rho}$$

in (3.E.21), one obtains that for $0 \le s \le \rho \le 1$ and $m = 1, \ldots, M$

$$\sum_{\{\mathbf{x}_i\}_{i=1}^M \setminus \{\mathbf{x}_m\}} \left(\prod_{i \ne m} q_{\mathbf{X}}(\mathbf{x}_i) \right) D_{\{\mathbf{x}_i\}_{i=1}^M}(m, G_n^m, s, \rho)$$
$$\le (M-1)^{\rho} \sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}_m) \left(\sum_{\mathbf{x}'} q_{\mathbf{X}}(\mathbf{x}') \left(\frac{p(\mathbf{y}|\mathbf{x}')}{p(\mathbf{y}|\mathbf{x}_m)} \right)^{\frac{s}{\rho}} \right)^{\rho}.$$

By averaging $D_{\{\mathbf{x}_i\}_{i=1}^M}(m, G_n^m, s, \rho)$ over the M codewords, we get that for every index m $(1 \le m \le M)$

$$\sum_{\{\mathbf{x}_i\}_{i=1}^M} \left(\prod_{i=1}^M q_{\mathbf{X}}(\mathbf{x}_i) \right) D_{\{\mathbf{x}_i\}_{i=1}^M}(m, G_n^m, s, \rho)$$
$$= \sum_{\mathbf{x}_m} q_{\mathbf{X}}(\mathbf{x}_m) \sum_{\{\mathbf{x}_i\}_{i=1}^M \setminus \{\mathbf{x}_m\}} \left(\prod_{i \ne m} q_{\mathbf{X}}(\mathbf{x}_i) \right) D_{\{\mathbf{x}_i\}_{i=1}^M}(m, G_n^m, s, \rho)$$
$$\le (M-1)^{\rho} \sum_{\mathbf{y}} \sum_{\mathbf{x}_m} q_{\mathbf{X}}(\mathbf{x}_m) p(\mathbf{y}|\mathbf{x}_m) \left(\sum_{\mathbf{x}'} q_{\mathbf{X}}(\mathbf{x}') \left(\frac{p(\mathbf{y}|\mathbf{x}')}{p(\mathbf{y}|\mathbf{x}_m)} \right)^{\frac{s}{\rho}} \right)^{\rho}$$
$$= (M-1)^{\rho} \sum_{\mathbf{y}} \left\{ \left(\sum_{\mathbf{x}} q_{\mathbf{X}}(\mathbf{x})\, p(\mathbf{y}|\mathbf{x})^{1-s} \right) \left(\sum_{\mathbf{x}'} q_{\mathbf{X}}(\mathbf{x}') p(\mathbf{y}|\mathbf{x}')^{\frac{s}{\rho}} \right)^{\rho} \right\}. \tag{3.E.22}$$

Since the right-hand side of (3.E.22) does not depend on the index m, then this bound also applies to the expectation of the quantity $\frac{1}{M} \sum_{m=1}^M D_{\{\mathbf{x}_i\}_{i=1}^M}(m, G_n^m, s, \rho)$. Therefore, there exists a block code for which the value of this quantity is not larger than the average over the considered ensemble, i.e.,

$$\frac{1}{M} \sum_{m=1}^M D_{\{\mathbf{x}_i\}_{i=1}^M}(m, G_n^m, s, \rho)$$
$$\le (M-1)^{\rho} \sum_{\mathbf{y}} \left\{ \left(\sum_{\mathbf{x}} q_{\mathbf{X}}(\mathbf{x})\, p(\mathbf{y}|\mathbf{x})^{1-s} \right) \left(\sum_{\mathbf{x}'} q_{\mathbf{X}}(\mathbf{x}') p(\mathbf{y}|\mathbf{x}')^{\frac{s}{\rho}} \right)^{\rho} \right\}. \tag{3.E.23}$$

From (3.8), (3.9) and (3.E.23), it follows that the above block code satisfies simultaneously

$$\begin{aligned}
P_{\mathrm{e}} &= \frac{1}{M}\sum_{m=1}^{M} P_{\mathrm{e}|m} \\
&\leq e^{nsT}\cdot\frac{1}{M}\sum_{m=1}^{M} D_{\{\mathbf{x}_i\}_{i=1}^{M}}(m, G_n^m, s, \rho) \\
&\leq e^{nsT}(M-1)^{\rho}\sum_{\mathbf{y}}\left\{\left(\sum_{\mathbf{x}} q_{\mathbf{X}}(\mathbf{x})\, p(\mathbf{y}|\mathbf{x})^{1-s}\right)\left(\sum_{\mathbf{x}'} q_{\mathbf{X}}(\mathbf{x}')p(\mathbf{y}|\mathbf{x}')^{\frac{s}{\rho}}\right)^{\rho}\right\} \\
&< e^{n(sT+\rho R)}\sum_{\mathbf{y}}\left\{\left(\sum_{\mathbf{x}} q_{\mathbf{X}}(\mathbf{x})\, p(\mathbf{y}|\mathbf{x})^{1-s}\right)\left(\sum_{\mathbf{x}'} q_{\mathbf{X}}(\mathbf{x}')p(\mathbf{y}|\mathbf{x}')^{\frac{s}{\rho}}\right)^{\rho}\right\} \\
&= e^{-n\left(E_0(s,\rho,q_X)-\rho R-sT\right)}
\end{aligned}$$

and

$$\begin{aligned}
P_{\mathrm{ue}} &< e^{n\left((s-1)T+\rho R\right)}\sum_{\mathbf{y}}\left\{\left(\sum_{\mathbf{x}} q_{\mathbf{X}}(\mathbf{x})\, p(\mathbf{y}|\mathbf{x})^{1-s}\right)\left(\sum_{\mathbf{x}'} q_{\mathbf{X}}(\mathbf{x}')p(\mathbf{y}|\mathbf{x}')^{\frac{s}{\rho}}\right)^{\rho}\right\} \\
&= e^{-n\left(E_0(s,\rho,q_X)-\rho R-(s-1)T\right)}
\end{aligned}$$

where the last two equalities follow from (3.14), and since the input distribution and the channel are assumed to be memoryless, i.e.,

$$p(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{n} p(y_i|x_i), \quad q_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{n} q_X(x_i).$$

The proof of Corollary 3.1 is completed by optimizing the bounds over the parameters ρ and s (where $0 \leq s \leq \rho \leq 1$) and the input distribution q_X. This gives the exponents E_1 and E_2 in (3.13) for the upper bounds on P_{e} and P_{ue}, respectively.

3.F Proof of Proposition 3.6

The bounds in Proposition 3.6 are derived from Proposition 3.5 as follows: setting

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

and

$$G_n^m(\mathbf{y}) = \prod_{i=1}^{n} g(y_i)$$

in (3.10), and relying on the useful rule for interchanging sum and product signs $\sum_{\mathbf{y}} \prod_{i=1}^{n} f(y_i) = \prod_{i=1}^{n} \sum_{y_i} f(y_i)$, one gets from (3.8) the RHS of (3.15) as an upper bound on $P_{\mathrm{e}|0}$. Since the considered block code is linear and the communication channel is memoryless and symmetric, the bound in (3.15) follows from the message independence property in Proposition 3.2. The derivation of the bound in (3.16) relies on (3.9) where it is first proved that for a linear block code whose transmission takes place over a memoryless symmetric channel, the resulting expression for $D_{\mathrm{B}}(m, G_n^m, s, \rho)$ is independent of m. To this end, let $\mathcal{T}$ be a mapping as defined in Definition 2.1, then for all $1 \leq i \leq n$

$$\begin{aligned}
&\sum_{m' \neq m} \sum_{y \in \mathcal{Y}} g(y)^{1-\frac{1}{\rho}} p(y|x_{m,i}) \left(\frac{p(y|x_{m',i})}{p(y|x_{m,i})} \right)^{\frac{s}{\rho}} \\
&= \sum_{m' \neq m} \sum_{y \in \mathcal{Y}} g(y)^{1-\frac{1}{\rho}} p(\mathcal{T}(y, -x_{m,i})|0) \left(\frac{p(\mathcal{T}(y, -x_{m,i})|x_{m',i} - x_{m,i})}{p(\mathcal{T}(y, -x_{m,i})|0)} \right)^{\frac{s}{\rho}} \\
&= \sum_{l \neq 0} \sum_{z \in \mathcal{Y}} g(y)^{1-\frac{1}{\rho}} p(z|0) \left(\frac{p(z|x_{l,i})}{p(z|0)} \right)^{\frac{s}{\rho}}.
\end{aligned}$$

As a result, it follows that for a memoryless and symmetric channel

$$\frac{1}{M} \sum_{m=1}^{M} D_{\mathrm{B}}(m, G_n^m, s, \rho) = D(g, s, \rho) \tag{3.F.24}$$

where $D(g, s, \rho)$ is introduced in (3.17). The proof of the upper bound on P_{ue} as given in (3.16) is completed by substituting (3.F.24) in (3.16).

3.G Proof of Proposition 3.7

Proof of the upper bound on the conditional error probability in (3.40)

Let Λ_m^{LR} designate the decision region in (3.6), then the conditional error probability is equal to

$$P_{\mathrm{e}|m} = \sum_{\mathbf{y} \notin \Lambda_m^{\mathrm{LR}}} p(\mathbf{y}|\mathbf{x}_m).$$

For $\mathbf{y} \notin \Lambda_m^{\mathrm{LR}}$, the decision rule in (3.6) implies that

$$\frac{p(\mathbf{y}|\mathbf{x}_m)}{p(\mathbf{y}|\mathbf{x}_{m_2})} < e^{nT}$$

where $\mathbf{x}_{m_2}$ is the second most probable codeword, and therefore

$$e^{nT} \sum_{m' \neq m} \frac{p(\mathbf{y}|\mathbf{x}_{m'})}{p(\mathbf{y}|\mathbf{x}_m)} > 1.$$

Let $s \geq 0$, then for $\mathbf{y} \notin \Lambda_m^{\mathrm{LR}}$

$$e^{nTs}\left(\sum_{m' \neq m} \frac{p(\mathbf{y}|\mathbf{x}_{m'})}{p(\mathbf{y}|\mathbf{x}_m)}\right)^s > 1$$

and the conditional block error probability satisfies

$$P_{\mathrm{e}|m} \leq e^{nsT} \sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}_m) \left(\sum_{m' \neq m} \frac{p(\mathbf{y}|\mathbf{x}_{m'})}{p(\mathbf{y}|\mathbf{x}_m)}\right)^s. \tag{3.G.25}$$

The bound in (3.40) follows from (3.G.25), using the arguments following (3.D.17).

Proof of the upper bound on the conditional undetected error probability in (3.41)

The conditional undetected error probability is given by

$$P_{\mathrm{ue}|m} = \sum_{\mathbf{y} \in \mathcal{L}} p(\mathbf{y}|\mathbf{x}_m)$$

where

$$\mathcal{L} \triangleq \left\{\mathbf{y} : \exists\, m' \neq m,\ p(\mathbf{y}|\mathbf{x}_{m'}) \geq e^{nT} p(\mathbf{y}|\mathbf{x}_{m_2'})\right\}$$

and $\mathbf{x}_{m_2'}$ is the second most probable codeword for $p(\mathbf{y}|\mathbf{x})$. Since $p(\mathbf{y}|\mathbf{x}_{m_2'}) \geq p(\mathbf{y}|\mathbf{x}_m)$, then

$$\mathcal{L} \subseteq \left\{\mathbf{y} : \exists\, m' \neq m,\ p(\mathbf{y}|\mathbf{x}_{m'}) \geq e^{nT} p(\mathbf{y}|\mathbf{x}_m)\right\}$$

and therefore

$$\begin{aligned}
\mathbf{y} \in \mathcal{L} \quad &\Rightarrow \quad \exists\, m' \neq m,\ \frac{p(\mathbf{y}|\mathbf{x}_{m'})}{p(\mathbf{y}|\mathbf{x}_m)} \cdot e^{-nT} \geq 1 \\
&\Rightarrow \quad e^{-nT} \sum_{m' \neq m} \frac{p(\mathbf{y}|\mathbf{x}_{m'})}{p(\mathbf{y}|\mathbf{x}_m)} \geq 1 \\
&\Rightarrow \quad \forall\, s \geq 0,\ e^{-nTs} \left(\sum_{m' \neq m} \frac{p(\mathbf{y}|\mathbf{x}_{m'})}{p(\mathbf{y}|\mathbf{x}_m)}\right)^s \geq 1
\end{aligned}$$

As a result, the conditional undetected block error probability satisfies, for all $s \geq 0$, the following upper bound:

$$P_{\mathrm{ue}|m} \leq e^{-nsT} \sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}_m) \left(\sum_{m' \neq m} \frac{p(\mathbf{y}|\mathbf{x}_{m'})}{p(\mathbf{y}|\mathbf{x}_m)}\right)^s.$$

The rest of the proof of (3.41) is, again, similar to the derivation following (3.D.17).

3.H Proof of Proposition 3.8

The derivation of the bounds in Proposition 3.8 is primarily identical to the analysis in [26] and [94, Section 3.2.4], for which the reader is referred for a complete treatment of the analysis under ML decoding. We assume a BPSK modulation over AWGN channel with energy E_{s} per transmitted coded symbol, and a white Gaussian noise with two-sided power spectral density of $\frac{N_0}{2}$. Hence, the received vector $\mathbf{y}$ satisfies

$$\mathbf{y} = \gamma \mathbf{x} + \mathbf{n} \tag{3.H.26}$$

where $\gamma \triangleq \sqrt{\frac{2E_{\mathrm{s}}}{N_0}}$, $\mathbf{x} \in \mathcal{C} \subseteq \{-1,+1\}^n$ is the transmitted codeword (with BPSK modulation), and $\mathbf{n}$ is a normal random vector with independent coordinates (all with zero mean and unit variance). Setting

$$E_{\mathrm{e}}(d) \triangleq \left\{ \mathbf{y} \in \mathcal{Y}^n : \ \frac{\max_{\mathbf{x} \in \mathcal{C}_d \setminus \{\mathbf{x}_0\}} p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y}|\mathbf{x}_0)} \cdot e^{nT} \geq 1 \right\}.$$

where $\mathcal{C}_d$ is the set of all codewords whose Hamming weight is d, and $\mathbf{x}_0$ is the all-zero codeword, it follows from (3.6) and the union bound that the conditional decoding error probability is upper bounded by

$$P_{\mathrm{e}|0} \leq \sum_{d=d_{\min}}^{n} \Pr\left(E_{\mathrm{e}}(d)\right) \tag{3.H.27}$$

where $d_{\min}$ denotes the minimal Hamming distance of $\mathcal{C}$. Consider the following inequality on the probability of an error event:

$$\Pr(E) \leq \Pr(E, \mathbf{y} \in \mathcal{R}) + \Pr(\mathbf{y} \notin \mathcal{R}) \tag{3.H.28}$$

where E denotes an error event, $\mathbf{y} \in \mathcal{Y}^n$ is the received vector, and $\mathcal{R} \subseteq \mathbf{Y}^n$. From (3.H.27) and (3.H.28), it follows that

$$P_{\mathrm{e}|0} \leq \sum_{d=d_{\min}}^{n} \Big(\Pr\big(E_{\mathrm{e}}(d), \mathbf{y} \in \mathcal{R}\big) + \Pr\big(\mathbf{y} \notin \mathcal{R}\big) \Big). \tag{3.H.29}$$

Using the union bound, we have

$$\begin{aligned} \Pr\big(E_{\mathrm{e}}(d), \mathbf{y} \in \mathcal{R}\big) &\leq \sum_{\mathbf{x} \in \mathcal{C}_d} \Pr\left(\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y}|\mathbf{x}_0)} \, e^{nT} \geq 1, \ \mathbf{y} \in \mathcal{R} \right) \\ &\overset{\text{(a)}}{=} \sum_{\mathbf{x} \in \mathcal{C}_d} \Pr\left(\langle \mathbf{y}, \mathbf{x} \rangle \geq \langle \mathbf{y}, \mathbf{x}_0 \rangle - \frac{nT}{\gamma}, \ \mathbf{y} \in \mathcal{R} \right) \end{aligned} \tag{3.H.30}$$

where equality (a) follows from (3.H.26), and $\langle \mathbf{x}, \mathbf{y} \rangle \triangleq \sum_{i=1}^{n} x_i y_i$ denotes the scalar multiplication of the vectors $\mathbf{x}$ and $\mathbf{y}$. Similarly to the derivation of bound in [26] (under ML decoding), we choose

$$\mathcal{R} \triangleq \left\{ \mathbf{y} : \ \|\mathbf{y} - \eta\gamma\mathbf{x}_0\|^2 \leq nr^2 \right\} \tag{3.H.31}$$

where η and r are arbitrary parameters which are subject to optimization. In addition, define

$$\begin{aligned} Z &\triangleq \langle \mathbf{y}, \mathbf{x} \rangle - \langle \mathbf{y}, \mathbf{x}_0 \rangle \\ W &\triangleq \|\mathbf{y} - \eta\gamma\mathbf{x}_0\|^2 - nr^2 \end{aligned}$$

then it follows from (3.H.30) and (3.H.31), using the Chernoff bound that

$$\Pr\big(E_{\mathrm{e}}(d),\ \mathbf{y} \in \mathcal{R}\big) + \Pr\big(\mathbf{y} \notin \mathcal{R}\big) \leq e^{\frac{tnT}{\gamma}} \left|\mathcal{C}_d\right| \mathsf{E}\Big[e^{tZ+uW}\Big] + \mathsf{E}\Big[e^{sW}\Big] \tag{3.H.32}$$

for all $t \geq 0$, $u \leq 0$, and $s \geq 0$. Evaluating the expectations in (3.H.32) and setting $t = \frac{\gamma}{2}(1 - 2u\eta)$, we have similarly to [26] and [94, Section 3.2.4]:

$$\begin{aligned} \Pr\big(E_{\mathrm{e}}(d),\ \mathbf{y} \in \mathcal{R}\big) + \Pr\big(\mathbf{y} \notin \mathcal{R}\big) \leq{}& e^{\frac{nT(1-ru\eta)}{2}} \left|\mathcal{C}_d\right| e^{-nur^2} \left(f_1\left(\gamma, u, \eta\right)\right)^{n-d} \left(f_2\left(\gamma, u, \eta\right)\right)^d \\ &+ e^{-nsr^2} \left(f_1\left(\gamma, s, \eta\right)\right)^n \end{aligned} \tag{3.H.33}$$

where

$$\begin{aligned} f_1\left(\gamma, \alpha, \eta\right) &\triangleq \frac{e^{\frac{(1-\eta)^2\gamma^2\alpha}{1-2\alpha}}}{\sqrt{1-2\alpha}} \\ f_2\left(\gamma, \alpha, \eta\right) &\triangleq \frac{e^{-\frac{\gamma^2(1-2\alpha\eta^2)}{2}}}{\sqrt{1-2\alpha}}, \quad \alpha < \frac{1}{2}. \end{aligned}$$

Optimizing the term e^{nr^2} on the right-hand side of (3.H.33), gives

$$\Pr\big(E_{\mathrm{e}}(d), \mathbf{y} \in \mathcal{R}\big) + \Pr\big(\mathbf{y} \notin \mathcal{R}\big) \leq 2^{h_2\left(\frac{s}{s-u}\right)} A^{-\frac{u}{s-u}} B^{\frac{s}{s-u}}, \quad 0 < s < \frac{1}{2},\ u \leq 0 \tag{3.H.34}$$

where

$$\begin{aligned} A &\triangleq \left(f_1\left(\gamma, s, \eta\right)\right)^n \\ B &\triangleq e^{\frac{nT(1-ru\eta)}{2}} \left|\mathcal{C}_d\right| \left(f_1\left(\gamma, u, \eta\right)\right)^{n-d} \left(f_2\left(\gamma, u, \eta\right)\right)^d \end{aligned}$$

and h_2 designates the binary entropy function on base 2. Using the change of variables

$$\begin{aligned} \rho &\triangleq \frac{s}{s-u} \\ \beta &\triangleq \rho(1-2u) \\ \xi &\triangleq \rho(1-2u\eta) \end{aligned}$$

where $0 \leq \rho \leq 1, 0 \leq \beta \leq 1$, and $\xi \geq 0$, the bound in (3.H.34) transforms to

$$\Pr\big(E_{\mathrm{e}}(d),\ \mathbf{y} \in \mathcal{R}\big) + \Pr\big(\mathbf{y} \notin \mathcal{R}\big) \leq 2^{h_2(\rho)} e^{-nE(E_{\mathrm{s}}/N_0, d/n, \beta, \rho, \xi) + \frac{nT\xi}{2}} \tag{3.H.35}$$

where

$$E(c,\delta,\beta,\rho,\xi)$$
$$\triangleq -\rho r_n(\delta) - \frac{\rho}{2}\ln\left(\frac{\rho}{\beta}\right) - \frac{1-\rho}{2}\ln\left(\frac{1-\rho}{1-\beta}\right) + c\left(1-(1-\delta)\frac{\xi^2}{\beta} - \frac{(1-\xi)^2}{1-\beta}\right).$$

The parameters ρ, β and ξ are optimized in [26], [94] such that the error exponent $E(c,\delta,\beta,\rho,\xi)$ is maximized[2] (note that the bound for $T=0$ coincides with the bound which refers to ML decoding), setting the optimal parameters yields the first argument in (3.46). The second term inside the minimization on the right-hand side of (3.46) follows from a union bound on the error probability

$$P_{\mathrm{e}} \leq \sum_{d=d_{\min}}^{n} \sum_{\mathbf{x}\in\mathcal{C}_d} \Pr\left(\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y}|\mathbf{x}_0)} e^{nT} \geq 1\right)$$

where for every codeword $\mathbf{x} \in \mathcal{C}_d$

$$\Pr\left(\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y}|\mathbf{x}_0)} e^{nT} \geq 1\right) = Q\left(\gamma\sqrt{d} - \frac{nT}{2\gamma\sqrt{d}}\right).$$

The derivation of the upper bound on the undetected error probability follows some similar arguments, and is therefore omitted.

3.I Proof of Proposition 3.9

The main ingredient for proving the DS2 bound on the block error probability under ML decoding (and also the well known random-coding bound) is that for a received vector $\mathbf{y}$ which is not included in the decision region Λ_m as given in (3.2), the following inequality holds:

$$1 \leq \left(\sum_{m'\neq m}\left(\frac{p(\mathbf{y}|\mathbf{x}_{m'})}{p(\mathbf{y}|\mathbf{x}_m)}\right)^{\lambda}\right)^{\rho}, \quad \lambda,\rho \geq 0. \tag{3.I.36}$$

When an error event under fixed-size (L) list decoding is considered, there exists L distinct codewords, all different from the transmitted codeword, whose a-posterior probability is larger than the one of the transmitted codeword. Hence, the sum on the right-hand side of (3.I.36) is divided by L. Specifically for a received vector $\mathbf{y}$

[2] It is possible to obtain the optimized ρ and ξ when maximizing the entire exponent $E(c,\delta,\beta,\rho,\xi) + \frac{T\xi}{2}$. To this end, ξ needs to be shifted by $-\frac{T}{2}$ and the optimal ρ remains without change. The parameter β is required to be numerically optimized over $0 \leq \beta \leq 1$. Nevertheless, the resulting bound gives only a marginal gain over the bound which maximizes $E(c,\delta,\beta,\rho,\xi)$ without the addition of $\frac{T\xi}{2}$.

that results in an error event, the following inequality is satisfied:

$$1 \leq \left(\frac{1}{L} \sum_{m' \neq m} \left(\frac{p(\mathbf{y}|\mathbf{x}_{m'})}{p(\mathbf{y}|\mathbf{x}_m)} \right)^{\lambda} \right)^{\rho}, \quad \lambda, \rho \geq 0 \tag{3.I.37}$$

Following the derivation of the DS2 bound in [94, p. 96] where the right-hand side of (3.I.36) is replaced with (3.I.37) leads to the derivation of the bound in Proposition 3.9. This derivation is repeated for the sake of completeness. For an arbitrarily chosen probability measure $\psi_n^m(\mathbf{y})$ it follows that:

$$\begin{aligned} P_{\mathrm{e}|m} &\leq \sum_{\mathbf{y}} \psi_n^m(\mathbf{y}) \big(\psi_n^m(\mathbf{y})\big)^{-1} p(\mathbf{y}|\mathbf{x}_m) \left(\frac{1}{L} \sum_{m' \neq m} \left(\frac{p(\mathbf{y}|\mathbf{x}_{m'})}{p(\mathbf{y}|\mathbf{x}_m)} \right)^{\lambda} \right)^{\rho} \\ &= \sum_{\mathbf{y}} \psi_n^m(\mathbf{y}) \left(\big(\psi_n^m(\mathbf{y})\big)^{-\frac{1}{\rho}} \big(p(\mathbf{y}|\mathbf{x}_m)\big)^{\frac{1}{\rho}} \frac{1}{L} \sum_{m' \neq m} \left(\frac{p(\mathbf{y}|\mathbf{x}_{m'})}{p(\mathbf{y}|\mathbf{x}_m)} \right)^{\lambda} \right)^{\rho} \\ &\leq \left(\sum_{m' \neq m} \sum_{\mathbf{y}} \big(\psi_n^m(\mathbf{y})\big)^{1-\frac{1}{\rho}} \big(p(\mathbf{y}|\mathbf{x}_m)\big)^{\frac{1}{\rho}} \frac{1}{L} \left(\frac{p(\mathbf{y}|\mathbf{x}_{m'})}{p(\mathbf{y}|\mathbf{x}_m)} \right)^{\lambda} \right)^{\rho} \end{aligned}$$

where the last inequality follows from Jensen's inequality. Plugging $\psi_n^m(\mathbf{y})$ as in (3.D.18) concludes the proof.

Chapter 4

Optimal Erasure and List Decoding Schemes of Convolutional Codes

Chapter Overview

A modified Viterbi Algorithm (VA) with erasures and list-decoding is introduced. This algorithm is shown to yield the optimal decoding rule of Forney with erasures and variable list-size (see Definition 3.1). For the case of decoding with erasures, the optimal algorithm is compared to the simple algorithm of Yamamoto and Itoh [120]. The comparison shows a remarkable similarity in simulated performance. The chapter is based on the following paper:

E. Hof, I. Sason, and S. Shamai (Shitz), "'Optimal generalized decoding of convolutional codes," *Proceedings of the Tenth International Symposium on Communication Theory and Applications*, pp. 6–10, Ambleside, UK, July 2009.

This chapter is structured as follows: Section 4.1 proposes a modification to the VA, and Section 4.2 presents some numerical results for the optimal decoding of convolutional codes with erasures.

4.1 Optimal generalized decoding of convolutional codes over memoryless channels

In this section, a modified VA is presented for optimal decoding of convolutional codes with erasures. In addition, it is proved that this modification coincides with the optimal decoding rule in (3.4).

Assuming that all codewords are transmitted with equal a-priori probability, the joint probabilities in (3.4) can be replaced with conditional probabilities, and the decoding regions in (3.4) are given by:

$$\Lambda_m = \left\{ \mathbf{y} \in \mathcal{Y}^N : \ \frac{\Pr(\mathbf{y}|\mathbf{x}_m)}{\sum_{m' \neq m} \Pr(\mathbf{y}|\mathbf{x}_{m'})} \geq e^{NT} \right\}. \tag{4.1}$$

The standard VA provides the ML decision and its corresponding likelihood metric for the case at hand. Consequently, it remains to evaluate the denominator in (4.1) which is involved in the specification of the decision regions in [41].

Remark 4.1 Since

$$\frac{\Pr(\mathbf{y}, \mathbf{x}_m)}{\sum_{m' \neq m} \Pr(\mathbf{y}, \mathbf{x}_{m'})} = \frac{\Pr(\mathbf{y}, \mathbf{x}_m)}{\Pr(\mathbf{y}) - \Pr(\mathbf{y}, \mathbf{x}_m)}$$

the denominator of the LHS of the inequality in (3.4) can also be evaluated using the forward part of the Bahl, Cocke, Jelinek, and Raviv (BCJR) algorithm [7].

A convolutional code $\mathcal{C}$ with k inputs and n outputs for every time unit, and of memory length m is considered. The information sequence $\mathbf{u} = (\mathbf{u}_1, \ldots \mathbf{u}_B)$, of length kB symbols, is encoded (followed by a termination sequence) to form the codeword $\mathbf{x} = (\mathbf{x}_1, \ldots \mathbf{x}_{B+m})$ of length $n(B+m)$ symbols. We assume a memoryless channel, and denote the received sequence by $\mathbf{y}$. Each encoding operation, where k new inputs are introduced and n coded symbol outputs are transmitted at every time unit, is considered as a single time step. Let the metric for each branch in the trellis graph of $\mathcal{C}$ be

$$\mu(\mathbf{y}_t|\mathbf{x}_t) \triangleq \ln\big(p(\mathbf{y}_t|\mathbf{x}_t)\big), \quad 1 \leq t \leq B+m \tag{4.2}$$

where $\mathbf{y}_t$ is the vector of n received samples at the decoder for the time step t, and $\mathbf{x}_t$ is the vector of coded symbols which corresponds to the considered branch at time t. In addition, we define the (cumulative) metric for each path in the trellis of $\mathcal{C}$ by

$$\mu(\mathbf{y}^t|\mathbf{x}^t) \triangleq \sum_{i=1}^{t} \mu(\mathbf{y}_i|\mathbf{x}_i)$$

where $\mathbf{y}^t = (\mathbf{y}_1, \ldots, \mathbf{y}_t)$ is the vector of nt received samples up to time step t, $\mathbf{x}^t = (\mathbf{x}_1, \ldots, \mathbf{x}_t)$ is the vector of nt coded symbols of the considered path, and the sum is taken over all the t branches of this path. The set of nodes at a given time step t, which correspond to the possible encoder states in this time step, is denoted by $\mathcal{V}(t)$. For each node v in the trellis, the set of branches entering v is denoted by $\mathcal{B}_v$. The originating node of a trellis branch b, is denoted by v_b^{-1}, and the vector of output coded symbols of b is denoted by $\mathbf{x}(b)$.

A detailed description of the proposed algorithm is provided in Fig. 4.1. For the sake of simplicity, the algorithm in Fig. 4.1 is provided for the particular case of decoding a terminated convolutional code with erasures ($T > 0$). Steps 1(a) and (b) in Fig. 4.1 form the initialization actions for the standard VA. Starting from time step $t = m$, there exists a single surviving path in the trellis for each state whose cumulative metric is updated and stored. While proceeding along the trellis, steps 2(a)-2(d) in Fig. 4.1 are the familiar add-compare-select steps of the standard VA; for each state, the surviving path metric is chosen according to the maximal accumulate metric. Steps 1(c), 2(e) and 2(f) in Fig. 4.1 are the introduced modification. These steps allow the recursive evaluation of the sum in the denominator of (4.1). After this recursive evaluation along the trellis, the surviving path is selected, and the information bits are reconstructed according to the generalized decision rule in (4.1); else, an erasure is declared.

The following theorem assures that the suggested algorithm coincides with Forney's generalized decoding rule, as given in Definition 3.1:

Theorem 4.1 Consider the decoding of a terminated convolutional code using the algorithm in Fig. 4.1. Assuming that $\mathbf{x}_m$ is the codeword which corresponds to the surviving path, then the generalized metric μ_{G} satisfies:

$$e^{\mu_{\mathrm{G}}} = \sum_{m' \neq m} \Pr(\mathbf{y}|\mathbf{x}_{m'}).$$

Proof: Let $\mathcal{K}(v)$ denote the set of all possible paths entering a node v in the trellis graph of $\mathcal{C}$, except for the surviving path for v. We prove by induction that the generalized metric $\mu_{\mathrm{G}}(v)$ evaluated at $v \in \mathcal{V}(t)$ satisfies

$$e^{\mu_{\mathrm{G}}(v)} = \sum_{k \in \mathcal{K}(v)} \Pr(\mathbf{y}^t|\mathbf{x}_k^t) \tag{4.3}$$

where $\mathbf{y}^t$ is the received vector up to time t (included), and $\mathbf{x}_k^t$ is the vector of the first nt symbols of the k-th codeword. First, we check that (4.3) follows for $t = m$ where each state $v \in \mathcal{V}(m)$ has a single entering path. Hence, the sum in (4.3) is void (i.e., $\mathcal{K}(v) = \emptyset$) which coincides with the setting $\mu_{\mathrm{G}}(v) = -\infty$ for all $v \in \mathcal{V}(m)$ (step 1(c) in Fig. 4.1). Assume by induction that (4.3) holds for $t = \tau - 1 \geq m$, and it is required to prove that (4.3) also holds for the next time step $t = \tau$. Let $\mathcal{K}_{\mathrm{s}}(v) \subseteq \mathcal{K}(v)$ denote all the paths in $\mathcal{K}(v)$ entering v via the same branch as the survivor. For $t = \tau$, consider the temporary result after step 2(e) in Fig 4.1, and assume that the algorithm is currently handling the state $v \in \mathcal{V}(\tau)$. Following the

1. For each state $v \in \mathcal{V}(m)$:
   (a) Set the single path entering v as the survivor $\mathbf{s}(v)$.
   (b) Evaluate the surviving entering path metric $\mu(v)$.
   (c) Set the generalized metric $\mu_{\mathrm{G}}(v) = -\infty$.
2. Iterations over $m+1 \le t \le L+m$. For each state $v \in \mathcal{V}(t)$ do:
   (a) Evaluate for each entering branch $b \in \mathbf{B}(v)$:
   $$\mu_b = \tilde{\mu}_b + \mu(v_b^{-1})$$
   where $\tilde{\mu}_b$ is the branch metric of b, and $\mu(v_b^{-1})$ is the path metric of the survivor at the source node v_b^{-1} of b.
   (b) Find the entering branch with the maximal path metric:
   $$b^* = \arg \max_{b \in \mathbf{B}(v)} \mu_b.$$
   (c) Set an updated survivor: $\mathbf{s}(v) = \big(\mathbf{s}(v_{b^*}^{-1}), \mathbf{x}(b^*)\big)$.
   (d) Set an updated survivor path metric: $\mu(v) = \mu_{b^*}$.
   (e) Evaluate a temporary generalized metric:
   $$\mu_{\mathrm{G}}(v) = \mu_{\mathrm{G}}(v_{b^*}^{-1}) + \tilde{\mu}_{b^*}$$
   where $\mu_{\mathrm{G}}(v_{b^*}^{-1})$ is the generalized metric evaluated at the previous node $v_{b^*}^{-1}$ of the survivor path, and $\tilde{\mu}_{b^*}$ is the branch metric of the last branch of the survivor path.
   (f) For each of the rest of the entering branches $b \in \mathbf{B}(v) \setminus \{b^*\}$ do:
      i. Evaluate
      $$\zeta \triangleq \tilde{\mu}_{\mathrm{b}} + \max\big(\mu_G(v_b^{-1}), \mu(v_b^{-1})\big) + \ln\Big(1 + \exp\big(-\big|\mu_G(v_b^{-1}) - \mu(v_b^{-1})\big|\big)\Big)$$
      where $\tilde{\mu}_b$ is the branch metric of b, $\mu_G(v_b^{-1})$ and $\mu(v_b^{-1})$ are the generalized and standard path metrics at the initial node v_b^{-1} of the branch b.
      ii. Update the generalized metric:
      $$\mu_{\mathrm{G}}(v) = \max\big(\mu_G(v), \zeta\big) + \ln\Big(1 + \exp\big(-|\mu_G(v) - \zeta|\big)\Big).$$
3. If $\mu - \mu_{\mathrm{G}} > n(B+m)T$, return the survivor in the single node in $\mathcal{V}(B+m)$. Else, return an erasure.

Figure 4.1: Modified VA for optimal generalized decoding (with erasures) of terminated convolutional codes.

induction assumption, the temporary value of the generalized metric $\mu_{\mathrm{G}}(v)$ satisfies

$$\begin{aligned} e^{\mu_{\mathrm{G}}(v)} &= e^{\tilde{\mu}_{b^*}} \cdot e^{\mu_{\mathrm{G}}(v_{b^*}^{-1})} \\ &\overset{\text{(a)}}{=} e^{\tilde{\mu}_{b^*}} \sum_{k \in \mathcal{K}(v_{b^*}^{-1})} \Pr(\mathbf{y}^{\tau-1} | \mathbf{x}_k^{\tau-1}) \\ &\overset{\text{(b)}}{=} \sum_{k \in \mathcal{K}_{\mathrm{s}}(v)} \Pr(\mathbf{y}^{\tau} | \mathbf{x}_k^{\tau}) \end{aligned} \tag{4.4}$$

where $\mathbf{y}^t$ is the received vector up to time step t, $\mathbf{x}_k^t$ is vector of the first nt symbols of the codeword corresponding to a path k in the trellis graph, b^* is the entering branch of the maximizing path metric in step 2(b) of the algorithm, $v_{b^*}^{-1}$ is the source node of b^*. Equality (a) follows by the induction assumption for $t = \tau - 1$, and equality (b) follows from the memoryless property of the channel and the definition of the branch metric in (4.2).

Next, let b be a branch which is handled by the algorithm in step 2(f.i), and denote by $\mathcal{K}_b(v) \subseteq \mathcal{K}(v)$ the set of all the paths in $\mathcal{K}(v)$ entering v via the branch b. After step 2(f.i) terminates, the variable ζ satisfies:

$$\begin{aligned} e^{\zeta} &\overset{\text{(a)}}{=} e^{\mu_{\mathrm{G}}(v_b^{-1}) + \tilde{\mu}_b} + e^{\mu(v_b^{-1}) + \tilde{\mu}_b} \\ &\overset{\text{(b)}}{=} \sum_{k \in \mathcal{K}_b(v)} \Pr(\mathbf{y}^{\tau} | \mathbf{x}_k^{\tau}) \end{aligned} \tag{4.5}$$

where $\mathbf{y}^\tau$ forms the received vector up to time step τ, and $\mathbf{x}_k^\tau$ forms the sequence of the first $n\tau$ symbols of the codeword corresponding to a path k in the trellis, equality (a) follows from the equality

$$\ln(e^a + e^b) = \max(a, b) + \ln\left(1 + e^{-|a-b|}\right) \tag{4.6}$$

and equality (b) follows from the induction assumption, using the same arguments leading to (4.4). Finally, from (4.4)-(4.6), the update in step 2(f.ii) guarantees that (4.3) follows for $t = \tau$. Hence, by induction (4.3) follows for all $t \geq m$. ■

Remark 4.2 The complexity of the proposed algorithm is linear in the block length B, and is exponential in the constraint length m of the code. This is the same complexity characteristics as in the case of the standard VA.

Remark 4.3 (On generalized decoding with variable-size list) Consider the problem of generalized decoding with a variable list-size according to the optimal decoding rule in (3.4) (with $T < 0$). According to the random coding analysis in [41] for low rates, the decoded list size is small (it typically includes one codeword);

however, for higher code rates, the decoded list is likely to increase exponentially with the block length. Consequently, when practical decoding is of interest, some fixed limit on the decoded list is set. The following two options are suggested: the first, is to apply the (parallel) list VA as in [95] with the evaluation of the generalized metric as applied in the algorithm stated in Fig. 4.1. At the final step, only the survivors satisfying the condition in step 3, are left in the decoded list. As long as the size of the decoded list of the optimal decoding rule in Definition 3.1 is below the predetermined size-limit, Theorem 4.1 assures that the suggested modification coincides with the optimal decoding rule in [41]. The second option is to apply the evaluation of the generalized metrics in a serial implementation of the list VA (see, e.g., [70], [95], [81]). The serial implementation of the list VA iteratively produces a sequence of probable codewords where each iteration produces the next most probable path in the trellis graph of the code. After each iteration, the generalized metric of the decoded path is checked to satisfy the condition in step 3 of the algorithm, and the iterations stop if the condition fails. This scheme iteratively produces the list of codewords according to the optimal decoding rule in Definition 3.1. Since an exponentially amount of iterations is not practical, the algorithm needs to be stopped after a predetermined upper limit on the number of possible iterations. The resulting decoded list equals to the list under the optimal decoding rule only if the size of the optimal list is not larger than the predetermined limit.

Remark 4.4 (On knowledge of channel state information) Let $\mathbf{x}$ and $\mathbf{y}$ be vectors of size N over the channel input and output alphabets, respectively. The path metric

$$\mu(\mathbf{y}|\mathbf{x}) \triangleq \ln\big(p(\mathbf{y}|\mathbf{x})\big).$$

may be replaced with an erroneous metric μ' which does not rely on the complete channel state information. Consequently, for some applications, the implementation of the VA does not require complete channel state information at the receiver. Take for example a BSC with a crossover probability p. For this case:

$$\mu(\mathbf{y}|\mathbf{x}) = d_{\mathrm{H}}(\mathbf{y},\mathbf{x}) \ln\left(\frac{p}{1-p}\right) + N\ln(1-p)$$

where $d_{\mathrm{H}}(\mathbf{y},\mathbf{x})$ is the Hamming distance between $\mathbf{x}$ and $\mathbf{y}$. Another example is the AWGN channel with energy E_{s} per transmitted symbol and a two-sided density noise power spectrum $N_0/2$, where we have:

$$\mu(\mathbf{y}|\mathbf{x}) = \frac{2E_{\mathrm{s}}}{N_0}\mathbf{y}^{\mathrm{T}}\mathbf{x} - \left(\frac{E_{\mathrm{s}}}{N_0}\left(\mathbf{y}^{\mathrm{T}}\mathbf{y}+N\right) - \frac{N}{2}\ln\frac{E_{\mathrm{s}}}{\pi N_0}\right).$$

A weakness of the proposed algorithm is that its proof of optimality according to Theorem 4.1 does not necessarily follow if the metric μ is replaced with an erroneous metric. The implication of this observation is that the proposed algorithm may require complete channel state information to guarantee its optimality according to Theorem 4.1.

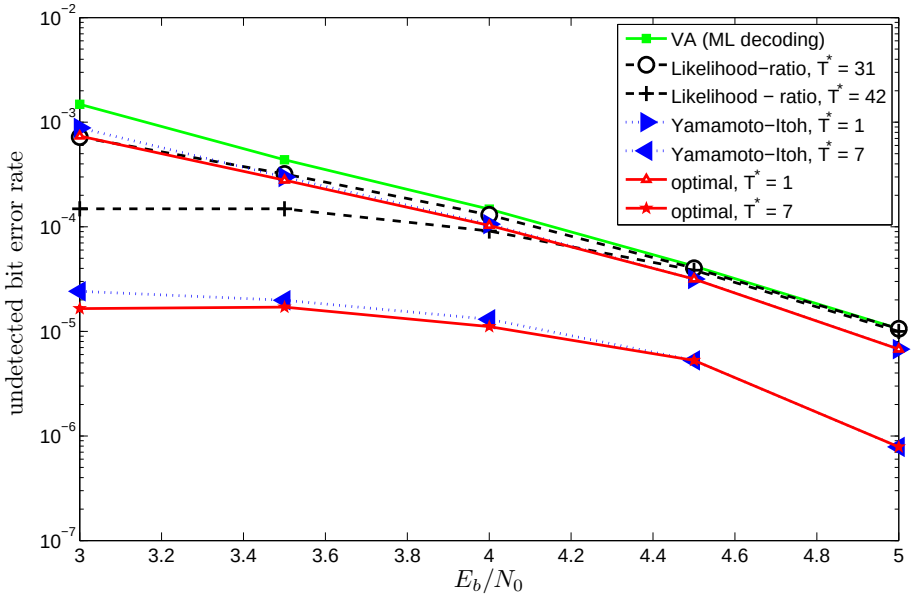
4.2 Examples

The performance of the (2,1,4) convolutional code with generator polynomials $g_1(D) = 1+D+D^3+D^4$ and $g_2(D) = 1+D^3+D^4$ is simulated under some generalized decoding algorithms with erasures. This code is used in the GSM Phase 2 system for the full-rate data traffic channel [1]. The results are provided for information sequence of 240 bits, with additional 4 bits of termination sequence (these bits are called "tail bits" in [1] where the same parameters are used). It is assumed that the transmission takes place over an AWGN channel with a binary phase shift keying (BPSK) modulation. Exact likelihood metrics are used in this simulation assuming complete channel state information at the decoder, i.e., the metric used in the simulation is $\mu(\mathbf{y}|\mathbf{x}) = 2\frac{E_s}{N_0}\mathbf{y}^{\mathrm{T}}\mathbf{x}$. Denote $T^* \triangleq n(B+m)T$, then the threshold e^{NT} in (3.4) and (3.6), is equal to e^{T^*}. In the following simulated results, the error performance for different values of the threshold parameter T^* are plotted. Note however, that when T^* is fixed with the applied metric, it follows that a different receiver is simulated for each SNR value. In Figure 4.2(a), the undetected bit error rates under the optimal generalized decoding algorithm in Figure 4.1 (based on the optimal decoding rule of Forney [41]), with $T^* = 1$ and 7, are provided. In addition, the bit error rate of the standard VA and the undetected bit error rate of the LR decoding in (3.6) with $T^* = 31$ and 42, and under the decoding algorithm of Yamamoto-Itho (this algorithm uses a threshold A (see [120, Section II]), the same T^* values of the optimal algorithm are used for A.) (YI) [120] , are provided for comparison. The corresponding block erasure rates for the simulated algorithms are provided in Figure 4.2(b). It is evident that the estimated performance of the optimal algorithm outperforms the one of the LR decoding rule. The undetected error performance of the optimal algorithm with $T^* = 1$, resembles the undetected bit error rates of the LR decoding rule with $T^* = 31$ and 42. However, the corresponding erasure rates under optimal decoding clearly outperform the suboptimal erasure rates under the LR decoding rule. Moreover, the optimal algorithm with $T^* = 7$, which results in similar erasure rates as the LR decoding rule with $T^* = 31$, whereas its undetected bit error rates clearly outperforms the undetected bit error rates under the LR decoding rule. Comparing

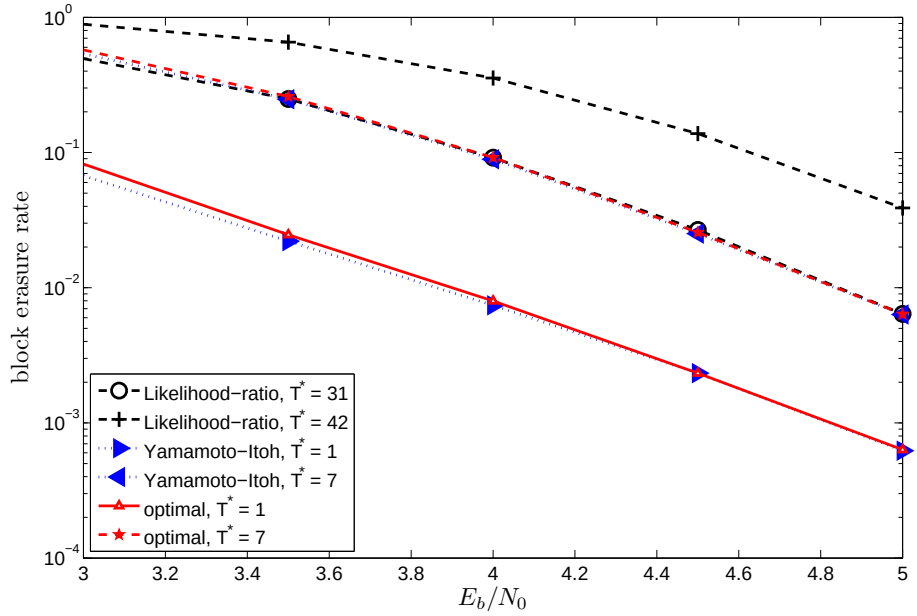
the simulated performance of the optimal algorithm with the YI decoding algorithm shows a remarkable improvement as compared to the LR decoding rule, and a good match with the simulated performance under the optimal decoding rule. For high SNR values, both decoding algorithms show the almost the same performance. For low SNR values, the gain of the optimal algorithm as compared with the YI algorithm is marginal. Take for example the results for $T^* = 7$ where both decoding algorithms have almost the same erasure rates, while only a slight improvement of the undetected bit error rate is observed for the optimal algorithm (in low SNR values).

4.3 Summary and Conclusions

An optimal algorithm is provided based on the generalized decision regions of Forney [41]. This algorithm allows for a practical generalized decoding of convolutional codes with erasures and variable list-sizes. The simulated performance of the proposed algorithm is compared with two suboptimal erasure decoding algorithms: the LR decoding rule in (3.6), and an algorithm by Yamamoto and Itoh (YI) [120]. The difference between the simulated performance of the optimal decoding algorithm and the YI algorithm is negligible. Moreover, the implementation of the YI algorithm is simpler and it yields a remarkable reduction in decoding complexity. The performance of the LR decoding rule, on the other hand, are substantially inferior.



(a)



(b)

Figure 4.2: Error performance of a (2,1,4) convolutional code under generalized decoding with erasures. Undetected bit error rates, and erasure rates, are provided in plots (a) and (b), respectively, under the optimal decoding in Figure 4.1, the LR decoding rule in (3.6), and for the Yamamoto-Itoh (YI) decoding algorithm [120]. The bit error rate under ML decoding (using the standard VA) is also provided. The results are provided for information sequence of 240 bits, with additional 4 bits of termination sequence

Chapter 5

Secrecy-Achieving Polar-Coding

Chapter Overview

A secrecy polar scheme is provided in this chapter for the two-user wire-tap channel model. A secret message needs to be transmitted reliably to a legitimate user. At the same time, this message must be kept secret from the eavesdropper. It is assumed that the marginal channel to the eavesdropper is physically degraded with respect to the marginal channel to the legitimate user. The proposed secrecy polar scheme for the degraded case is based on transmitting random bits on the 'good bits' of the degraded eavesdropper channel. These random bits are independent of the secret message. The 'good bits' for the degraded eavesdropper channel are also 'good' for the legitimate user. Consequently, these random bits can be decoded reliably at the legitimate user. The rest of the 'good' bits for the legitimate user are dedicated for the secret message.

Transmitting random bits on the 'good bits' of the eavesdropper, all the possible information rates that can be detected by the eavesdropper are exhausted. Otherwise, the standard channel capacity could have been beaten. Thus the 'good bits' associated with the secret message for the legitimate channel, must be perfectly secret (at least in the weak sense). Note that this result is satisfied immaterial of whether the eavesdropper adheres to successive decoding or to optimal decoding (as otherwise, its capacity could have been beaten). The chapter is based on the following paper:

E. Hof and S. Shamai (Shitz), "Secrecy-Achieving Polar-Coding," submitted to the *IEEE Trans. on Information Theory*, May 2010. This work is presented in part in the *2010 IEEE Information Theory Workshop (ITW 2010),* Dublin, Irland, September 2010 (Invited talk).

Additional independent works on this subject are provided in [3] [66] [75].

This chapter is structured as follows: In Section 6.1 preliminary introduction is provided. In Section 6.1.1 the wire-tap communication model is introduced in addition to some basic definitions and results in information-theoretic security. Polar codes are introduced in Section 5.1.2. The polar secrecy scheme is detailed and studied in Section 6.3. A conjecture on possible polarization properties is stated in Section 5.3, along with a resulting adaptation of the polar secrecy scheme for non-degraded wiretap channels. A list of possible further generalizations is provided in Section 5.4.

5.1 Preliminaries

5.1.1 The Wire-Tap Communication Model

We consider the communication model in Figure 5.1. A coded system is presented which transmits a confidential message U to a legitimate user. The message U is chosen uniformly from a set of size M. Next, the message is encoded to a codeword $\mathbf{X}$ with a blocklength n over an alphabet $\mathcal{X}$. The resulting code-rate is $R = \frac{1}{n}\log M$. The codeword $\mathbf{X}$ is transmitted over a DMC P, with an input alphabet $\mathcal{X}$, and output alphabets $\mathcal{Y}$ and $\mathcal{Z}$. Let $P(\mathbf{y}, \mathbf{z}|\mathbf{x})$ denote the probability of receiving the vectors $\mathbf{y} \in \mathcal{Y}^n$, and $\mathbf{z} \in \mathcal{Z}^n$, at the legitimate user and the eavesdropper, respectively, given that a codeword $\mathbf{x} \in \mathcal{X}^n$ is transmitted. Based on the assumption that the channel is memoryless, it follows that

$$P(\mathbf{y}, \mathbf{z}|\mathbf{x}) = \prod_{k=1}^{n} P(y_k, z_k|x_k)$$

where (with some abuse of notation) $P(y, z|x)$ denotes the probability of receiving the symbols $y \in \mathcal{Y}$ and $z \in \mathcal{Z}$, at the legitimate user and the eavesdropper, respectively, given that the symbol $x \in \mathcal{X}$ is transmitted. Moreover, let $G(y|x)$ and $Q(z|x)$ denote the marginal probabilities for receiving the symbols $y \in \mathcal{Y}$ and $z \in \mathcal{Z}$, at the legitimate user and the eavesdropper, respectively, given that the symbol $x \in \mathcal{X}$ is transmitted. Both $G(y|x)$ and $Q(z|x)$ are transition probability laws of DMCs, called the marginal channels of the legitimate user and the eavesdropper, respectively. In addition, the probability to receive the symbol $z \in \mathcal{Z}$ at the eavesdropper, given that the symbol $y \in \mathcal{Y}$ is received at the legitimate user is denoted by $D(z|y)$.

The channel output vectors $\mathbf{Y}$ and $\mathbf{Z}$, both of length n, are received by the legitimate user and the eavesdropper, respectively. The legitimate user decodes the

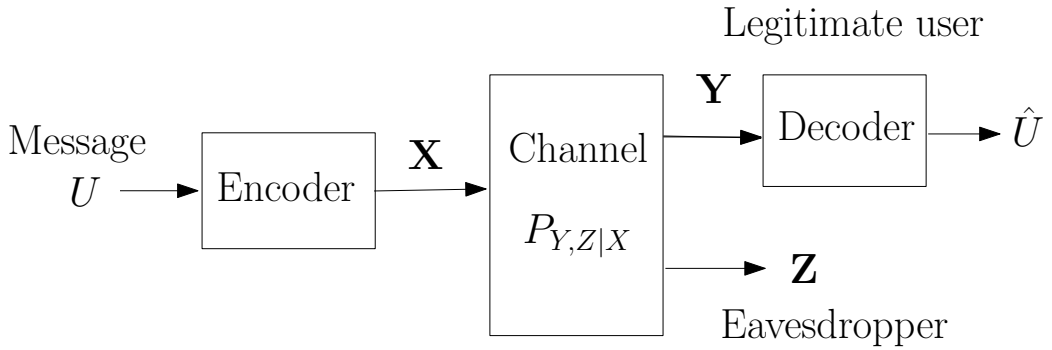


Figure 5.1: A wire-tap communication model.

received vector $\mathbf{Y}$ resulting in the decoded message $\hat{U}$. The objectives of the considered coding system is to obtain both secure and reliable communication. These objectives are to be accomplished simultaneously using a single codebook $\mathcal{C}_n$. The reliability of the system is measured via the average error probability $P_{\mathrm{e}}(\mathcal{C}_n)$ of the decoded message

$$P_{\mathrm{e}}(\mathcal{C}_n) = \frac{1}{M}\sum_{m=1}^{M} \Pr\left(\hat{U} \neq m |\ U = m\right).$$

Note that the error probability depends on the blocklength of the coded message. The level of security is measured by the equivocation rate

$$R_{\mathrm{e}}(\mathcal{C}_n) \triangleq \frac{1}{n} H(U|\mathbf{Z}) \tag{5.1}$$

where $H(U|\mathbf{Z})$ denotes the conditional entropy of the transmitted message U, given the received vector $\mathbf{Z}$ at the eavesdropper.

Definition 5.1 (Achievable rate-equivocation pair) A rate-equivocation pair (R, R_{e}) is achievable if there exists a code sequence $\{\mathcal{C}_n\}$ of block length n and rate R such that

$$\lim_{n\to\infty} P_{\mathrm{e}}(\mathcal{C}_n) = 0$$
$$R_{\mathrm{e}} \leq \lim_{n\to\infty} R_{\mathrm{e}}(\mathcal{C}_n).$$

Remark 5.1 (On strong and weak notions of secrecy) The current discussion considers normalized entropies to measure the level of security (see the definition of equivocation rate in (5.1)). Therefore, the achieved secrecy notion is referred to as *weak* secrecy. The *strong* notion of secrecy considers the unnormalized mutual information between the confidential message and the received vector at the eavesdropper receiver. Strong secrecy guarantees secrecy in the weak sense while the opposite direction does not follow.

Definition 5.2 (Secrecy capacity) The secrecy capacity C_{s} is the supremum of all the rates R, such that the pair (R, R) is an achievable rate-equivocation pair.

Theorem 5.1 (The secrecy capacity of the wire-tap channel [69]) The secrecy capacity C_{s} of the wire-tap channel satisfies:

$$C_{\mathrm{s}} = \max_{P_{UX} P_{YZ|X}} \big(I(U;Y) - I(U;Z)\big)$$

where U is an auxiliary random variable over the alphabet $\mathcal{U}$, satisfying

1. Markov relationship: $U \to X \to (Y, Z)$ is a Markov chain.
2. Bounded cardinality: $|\mathcal{U}| \leq |\mathcal{X}| + 1$.

Binary-input symmetric wire-tap channels are considered in this chapter.

Definition 5.3 (Symmetric binary input channels) A DMC with a transition probability p, binary-input alphabet $\mathcal{X}$, and an output alphabet $\mathcal{Y}$ is said to be symmetric if there exists a permutation π over $\mathcal{Y}$ such that

1. The inverse permutation π^{-1} is equal to π, i.e.,

   $$\pi^{-1}(y) = \pi(y)$$

   for all $y \in \mathcal{Y}$.

2. The transition probability p satisfies

   $$p(y|0) = p(\pi(y)|1)$$

   for all $y \in \mathcal{Y}$.

Definition 5.4 (Symmetric binary-input wire-tap channels) A binary input discrete memoryless wire-tap channel is symmetric if both of its marginal channels are symmetric.

The particular case of physically degraded channels is studied in this chapter.

Definition 5.5 (Physically degraded channels) Let P be a wire-tap channel with an input alphabet $\mathcal{X}$ and output alphabets $\mathcal{Y}$ and $\mathcal{Z}$, at the legitimate and eavesdropper, respectively. Then, P is said to be physically degraded if

$$P(y, z|x) = G(y|x)D(z|y) \tag{5.2}$$

for all $x \in \mathcal{X}$, $y \in \mathcal{Y}$, and $z \in \mathcal{Z}$.

The following Theorem characterizes the secrecy capacity of a binary-input, memoryless, symmetric and degraded wire-tap channel:

Theorem 5.2 ([69]) Let P be a binary-input, memoryless, symmetric, and degraded wire-tap channel. Denote by $G_{Y|X}$ and $Q_{Z|X}$ the marginal channels to the legitimate user and the eavesdropper, respectively. Then, the secrecy capacity C_{s} is given by

$$C_{\mathrm{s}}(P) = C(G_{Y|X}) - C(Q_{Z|X})$$

where $C(G_{Y|X})$ and $C(Q_{Z|X})$ are the channel capacities of the marginal channel $G_{Y|X}$ and $Q_{Z|X}$, respectively.

Remark 5.2 (On the entire rate-equivocation region) Theorem 5.2 is a particular case of the rate-equivocation region of less-noisy channels (which is on its own a particular case of the rate-equivocation region of the wire-tap channel). Under the notation in Theorem 5.1, if $I(U;Y) \geq I(U;Z)$ for every U satisfying the Markov relationship in Theorem 5.1, then the channel to the legitimate receiver is said to be *less noisy* than the eavesdropper (the degradation assumption in (5.2) satisfies the less noisy condition). It can be shown for the case of less-noisy wire-tap channels, that the rate-equivocation region is given by

$$\bigcup_{P_X P_{YZ|X}} \left\{ (R, R_{\mathrm{e}}) : \begin{array}{l} 0 \leq R \leq I(X;Y) \\ 0 \leq R_{\mathrm{e}} \leq R \\ R_{\mathrm{e}} \leq I(X;Y) - I(X;Z) \end{array} \right\}.$$

For further details and proof see [69] and references therein. In the particular case of binary-input, memoryless symmetric and degraded wire-tap channels as in Theorem 5.2, the rate-equivocation region is therefore given by

$$\left\{ (R, R_{\mathrm{e}}) : \begin{array}{l} 0 \leq R \leq C(G_{Y|X}) \\ 0 \leq R_{\mathrm{e}} \leq R \\ R_{\mathrm{e}} \leq C(G_{Y|X}) - C(Q_{Z|X}) \end{array} \right\}. \tag{5.3}$$

5.1.2 Polar Codes

This preliminary section offers a short summary of the basic definitions and results in [4], [6], that are essential to the presentation of the results in Section 6.3.

Let p be a transition probability function of a DMC with a binary input-alphabet $\mathcal{X} = \{0, 1\}$ and an output alphabet $\mathcal{Y}$. The operation of the channel on vectors is

also denoted by p, that is for $\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$, and $\mathbf{y} = (y_1, \ldots, y_n) \in \mathcal{Y}^n$, the block transition probability is given by

$$p(\mathbf{y}|\mathbf{x}) = \prod_{l=1}^{n} p(y_l|x_l).$$

Polar codes are defined in [4] using the following recursive construction. At the first step, two independent copies of p are combined to form a new channel p_2 over an input alphabet $\mathcal{X}^2$ and output alphabet $\mathcal{Y}^2$. The transition probability function of the combined channel is given by

$$p_2(y_1, y_2|w_1, w_2) = p(y_1|w_1 + w_2)p(y_2|w_2) \tag{5.4}$$

for all $y_1, y_2 \in \mathcal{Y}$, and $w_1, w_2 \in \mathcal{X}$, where the addition operation is carried modulo 2. At the i-th step of the construction, the transition probability function p_n, for an integral power of 2, $n = 2^i$, is defined for a combined channel with an input alphabet $\mathcal{X}^n$ and an output alphabet $\mathcal{Y}^n$. The recursive definition of p_n is based on two independent copies of the channel $p_{\frac{n}{2}}$ defined at the previous step $(i-1)$. The channel $p_{\frac{n}{2}}$ has an input alphabet $\mathcal{X}^{\frac{n}{2}}$ and an output alphabet $\mathcal{Y}^{\frac{n}{2}}$. The construction of the channel p_n includes the following steps:

1. An input vector $\mathbf{w} = (w_1, \ldots, w_n) \in \mathcal{X}^n$ is first transformed to a vector $\mathbf{s} = (s_1, \ldots, s_n) \in \mathcal{X}^n$ where

   $$s_{2k-1} = w_{2k-1} + w_{2k}$$

   and

   $$s_{2k} = w_{2k}, \quad 1 \leq k \leq \frac{n}{2}$$

   where the addition is carried modulo 2.

2. The vector $\mathbf{s}$ is transformed into a vector $\mathbf{v} \in \mathcal{X}^n$ where

   $$\mathbf{v} = (s_1, s_3, \ldots, s_{n-1}, s_2, s_4, \ldots, s_n).$$

   i.e., the first $\frac{n}{2}$ elements of $\mathbf{v}$, $v_1, \ldots, v_{\frac{n}{2}}$, equal the elements in $\mathbf{s}$ with odd indices, and the other $\frac{n}{2}$ elements of $\mathbf{v}$, $v_{\frac{n}{2}+1}, \ldots, v_n$, are equal the elements of $\mathbf{s}$ with even indices. This operation is called a reverse shuffle operation and can be described by the linear transformation

   $$\mathbf{v} = \mathbf{s}R_n$$

   where R_n is an $n \times n$ matrix, called the reverse shuffle operator.

3. $p_n(\mathbf{y}|\mathbf{w})$ is given by

$$p_n(\mathbf{y}|\mathbf{w}) = p_{\frac{n}{2}}\left(y_1, y_2, \ldots, y_{\frac{n}{2}}|v_1, v_2, \ldots, v_{\frac{n}{2}}\right) \cdot p_{\frac{n}{2}}\left(y_{\frac{n}{2}+1}, y_{\frac{n}{2}+2}, \ldots, y_n|v_{\frac{n}{2}+1}, v_{\frac{n}{2}+2}, \ldots, v_n\right). \tag{5.5}$$

The recursive channel-synthesizing operation of p_n is referred to as channel combining, and the channel p_n is referred to as the combined channel. Note that all block lengths n are assumed to be integral powers of 2.

The recursive construction of p_n can be equivalently defined by using a linear encoding operation. Let

$$F = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

and define the following recursive construction of the $n \times n$ matrices G_n:

$$\begin{aligned} G_1 =& I_1 \\ G_n =& \left(I_{\frac{n}{2}} \otimes F\right) R_n \left(I_2 \otimes G_{\frac{n}{2}}\right) \end{aligned} \tag{5.6}$$

where I_l is the $l \times l$ identity matrix and $\otimes$ denotes the Kronecker product for matrices. The matrix G_n is refereed to as the polar generator matrix of size n.

Proposition 5.1 ([4]) Let p be a DMC, and let p_n be the combined channel with a block length n. Then,

$$p_n(\mathbf{y}|\mathbf{w}) = p(\mathbf{y}|\mathbf{w}G_n) \tag{5.7}$$

for all $\mathbf{y} \in \mathcal{Y}^n$ and $\mathbf{w} \in \mathcal{X}^n$, where p_n is the combined channel in (5.5) and G_n is the $n \times n$ matrix defined in (5.6).

Denote by $[n] \triangleq \{1, 2, \ldots, n\}$, and let $\mathcal{A}_n \subseteq [n]$. In addition, denote by $\mathcal{A}_n^{\mathrm{c}}$ the complementary set of $\mathcal{A}_n$, that is $\mathcal{A}_n^{\mathrm{c}} = [n] \setminus \mathcal{A}_n$. Given a set $\mathcal{A}_n$, a class of coset codes with a common code-rate $\frac{1}{n}|\mathcal{A}_n|$ are formed. Over the indices specified by $\mathcal{A}_n$, the components of the input vector $\mathbf{w}$ are set according to the information bit vector. The rest of the bits of $\mathbf{w}$ are predetermined and fixed according to the particular code design. By setting both the set $\mathcal{A}_n$ and the components of $\mathbf{w}$ specified by $\mathcal{A}_n^{\mathrm{c}}$, a particular coset code is defined. This code can be shown to be a block coset code. The set $\mathcal{A}_n$ is referred as the information set. Polar codes are constructed by a specific choice of the information set $\mathcal{A}_n$. Moreover, the choice of the information set is tailored to the specific channel over which the communication takes place.

A coset code is defined by using a linear block code and a coset vector. Let G be a generator matrix for a binary (n, k) linear block code with block length n and

dimension k. In addition, let $\mathbf{c} \in \mathcal{X}^n$ be a binary vector. Then, the coset block code $\mathbf{C}(G, \mathbf{c})$ is defined by

$$\mathcal{C}(G, \mathbf{c}) \triangleq \left\{\mathbf{x} : \ \mathbf{x} = \mathbf{u}G + \mathbf{c}, \ \mathbf{u} \in \mathcal{X}^k\right\}. \tag{5.8}$$

Denote by $G_n(\mathcal{A}_n)$ the $|\mathcal{A}_n| \times n$ sub-matrix of G_n, defined by the rows of G_n whose indices are in $\mathcal{A}_n$. Similarly, the matrix $G_n(\mathcal{A}_n^{\mathrm{c}})$ denotes the $|\mathcal{A}_n^{\mathrm{c}}| \times n$ sub-matrix of G_n formed by the remaining rows of G_n. For each choice of $\mathcal{A}_n$ and an arbitrary $n-k$ binary vector $\mathbf{b} \in \mathbf{X}^{n-k}$, $k = |\mathcal{A}_n|$, a coset code $\mathcal{C}$ is defined according to

$$\mathcal{C} = \mathcal{C}\big(G_n\left(\mathcal{A}_n\right), \mathbf{b}G_n\left(\mathcal{A}_n^{\mathrm{c}}\right)\big). \tag{5.9}$$

This coset coding construction coincides with the recursive construction in (5.6) and (5.7). Specifically, by proper choice of $\mathbf{w}$, $\mathbf{x} = \mathbf{w}G_n$. To see this, plug the information vector $\mathbf{u}$ in the information indices, specified by $\mathcal{A}_n$, of the input vector $\mathbf{w}$ to the recursive construction. In addition, plug the vector $\mathbf{b}$ in the rest of the components of $\mathbf{w}$.

Channel splitting is another important operation that is introduced in [4] for polar coding. The split channels $\{p_n^{(l)}\}_{l=1}^n$, with a binary input alphabet $\mathcal{X}$ and output alphabets $\mathcal{Y}^n \times \mathcal{X}^{l-1}$, $1 \le l \le n$, are defined according to:

$$p_n^{(l)}(\mathbf{y}, \mathbf{w}|x) = \frac{1}{2^{n-1}} \sum_{\mathbf{c} \in \mathcal{X}^{n-l}} p_n\big(\mathbf{y}|(\mathbf{w}, x, \mathbf{c})\big) \tag{5.10}$$

where $\mathbf{y} \in \mathcal{Y}^n$, $\mathbf{w} \in \mathcal{X}^{l-1}$, and $x \in \mathcal{X}$. The channel synthesizing operation in (5.10) is referred to as channel splitting operation. The Bhattacharyya parameter of $p_n^{(l)}$ is denoted by:

$$B(p_n^{(l)}) \triangleq \sum_{\mathbf{y} \in \mathcal{Y}^n} \sum_{\mathbf{w} \in \mathcal{X}^{l-1}} \sqrt{p_n^{(l)}(\mathbf{y}, \mathbf{w}|0) p_n^{(k)}(\mathbf{y}, \mathbf{w}|1)}. \tag{5.11}$$

The construction of the sequence of sets of split channels $\{p_n^{(l)}(\mathbf{y}, \mathbf{w}|x)\}_{l=1}^n$, $n = 2^i$, $i \in \mathbb{N}$, in (5.10) can be described using the following alternative recursion:

Proposition 5.2 ([4]) For all $i > 0$, $1 \le l \le 2^i$,

$$p_{2^{i+1}}^{(2l-1)}\big((\mathbf{y}^{(1)}, \mathbf{y}^{(2)}), \mathbf{w}|w_1\big) = \sum_{w \in \mathcal{X}} \frac{1}{2} p_{2^i}^{(l)}\big(\mathbf{y}^{(1)}, g(\mathbf{w})|w_1 + w\big) p_{2^i}^{(l)}\big(\mathbf{y}^{(2)}, e(\mathbf{w})|w\big) \tag{5.12}$$

$$p_{2^{i+1}}^{(2l)}\big((\mathbf{y}^{(1)}, \mathbf{y}^{(2)}), (\mathbf{w}, w_1)|w_2\big) = \frac{1}{2} p_{2^i}^{(l)}\big(\mathbf{y}^{(1)}, g(\mathbf{w})|w_1 + w_2\big) p_{2^i}^{(l)}\big(\mathbf{y}^{(2)}, e(\mathbf{w})|w_2\big) \tag{5.13}$$

where $\mathbf{y}^{(1)}, \mathbf{y}^{(2)} \in \mathcal{Y}^{2^i}$, $\mathbf{w} = (w_1, \ldots, w_{2l-2}) \in \mathcal{X}^{2l-2}$, $w_1, w_2 \in \mathcal{X}$, the addition operation is carried modulo 2 and $\mathbf{g} = (g_1, \ldots, g_{l-1}) = g(\mathbf{w})$ is a vector in $\mathcal{X}^{l-1}$ defined according to

$$g_j = w_{2j-1} + w_{2j}, \quad 1 \le j \le l-1 \tag{5.14}$$

and

$$e(\mathbf{w}) = (w_2, w_4, \ldots, w_{2l-2}) \tag{5.15}$$

is the vector in $\mathcal{X}^{l-1}$ comprises from the components of $\mathbf{x}$ with even indices.

The importance of channel splitting is in its role in the successive cancellation decoding procedure that is provided in [4]. The error performance analysis of this decoding procedure relies on the following two results:

Theorem 5.3 ([6]) Let p be a binary-input symmetric DMC with capacity $C(p)$, and fix an arbitrary rate $R < C(p)$ and a positive constant $\beta < \frac{1}{2}$. Then, there exists a sequence of information sets $\mathcal{A}_n \subset [n]$, where $n = 2^i$, $i \in \mathbb{N}$, such that for large enough blocklengths n the following properties are satisfied:

1. Rate:
$$|\mathcal{A}_n| \geq nR.$$

2. Performance: The Bhattacharyya parameters in (5.11) satisfy
$$B(p_n^{(l)}) \leq 2^{-n^\beta}$$
for every $l \in \mathcal{A}_n$.

Proposition 5.3 ([4]) Assume that the vector $\mathbf{w} = (w_1, \ldots, w_n) \in \mathcal{X}^n$ is encoded via the considered recursive construction in (5.7), and is transmitted over a memoryless and symmetric DMC channel p with a binary-input alphabet $\mathcal{X}$ and an output alphabet $\mathcal{Y}$. Define the event

$$\mathcal{E}_l(p) \triangleq \big\{p_n^{(l)}(\mathbf{y}, \mathbf{w}^{(l-1)}|w_l) \leq p_n^{(l)}(\mathbf{y}, \mathbf{w}^{(l-1)}|w_l + 1)\big\} \tag{5.16}$$

where $\mathbf{y} \in \mathcal{Y}^n$ is the received vector, $\mathbf{w}^{(l-1)} = (w_1, \ldots, w_{l-1})$ is the vector comprises of the first $l-1$ bits of $\mathbf{w}$, $p_n^{(l)}$ is the split channel in (5.10) and the addition is carried modulo 2. Then, the event $\mathcal{E}_l$ is independent of the actual input vector $\mathbf{w}$ and

$$\Pr\big(\mathcal{E}_l(p)\big) \leq B\big(p_n^{(l)}\big)$$

where $B(p_n^{(l)})$ is the Bhattacharyya parameter in (5.11).

5.2 The Proposed Scheme

5.2.1 Polar Coding for Degraded Wire-Tap Channels

Coset Block Codes

A polar coding scheme is defined for the wire-tap channel. The proposed scheme is defined using the notion of coset block codes, based on the polar generator matrix G_n introduced in Section 5.1.2. For a given block length $n = 2^i$, $i \in \mathbb{N}$, let $\mathcal{A}_n$ be an arbitrary subset of $[n]$ of size k. In addition, let $\mathcal{N}_n$ be an additional arbitrary subset of $\mathcal{A}_n^{\mathrm{c}}$, of size k^*, and let $\mathbf{b}_n \in \mathcal{X}^{n-k-k^*}$ be a length $n-k-k^*$ binary vector. Denote by $\mathcal{B}_n$ the set of remaining indices in $\mathcal{A}_n^{\mathrm{c}}$, that is

$$\mathcal{B}_n \triangleq \mathcal{A}_n^{\mathrm{c}} \setminus \mathcal{N}_n. \tag{5.17}$$

The sets $\mathcal{A}_n$, $\mathcal{B}_n$, and $\mathcal{N}_n$, the polar generator matrix G_n and the vector $\mathbf{b}_n$ are all known to both the legitimate user and the eavesdropper.

Let $\mathbf{u} \in \mathcal{X}^k$ be a confidential information bit vector that needs to be transmitted to the legitimate user. The operation of the proposed secrecy scheme is described as follows:

1. A binary vector $\mathbf{b}_n^* \in \mathcal{X}^{k^*}$ is chosen uniformly at random.

2. The coset block code $\mathcal{C}_n^*$ is chosen according to

$$\mathcal{C}_n^* = \mathcal{C}\big(G_n(\mathcal{A}_n), \mathbf{b}_n G_n(\mathcal{B}_n) + \mathbf{b}_n^* G_n(\mathcal{N}_n)\big). \tag{5.18}$$

3. The information vector $\mathbf{u}$ is encoded into a codeword $\mathbf{x}$ using the coset block code $\mathcal{C}_n^*$. That is,

$$\mathbf{x} = \mathbf{u}G_n(\mathcal{A}_n) + \mathbf{b}_n G_n(\mathcal{B}_n) + \mathbf{b}_n^* G_n(\mathcal{N}_n) \tag{5.19}$$

   and it is transmitted over the wire-tap channel.

As the complexity of constructing a random vector can be assumed to be $\mathcal{O}(n)$, then the encoding complexity of the proposed scheme equals the encoding complexity of the single-user polar encoding in [4], which is $\mathcal{O}(n \log n)$.

For given sets $\mathcal{A}_n$ and $\mathcal{N}_n$, and a vector $\mathbf{b}_n$, the resulting coding scheme is denoted by $\mathcal{C}_n(\mathcal{A}_n, \mathcal{N}_n, \mathbf{b}_n)$. Since symmetric channels are considered, the performance of the provided scheme is shown in the following to be independent of the actual choice of $\mathbf{b}_n$. Consequently, the suggested coding scheme is denoted by $\mathcal{C}_n(\mathcal{A}_n, \mathcal{N}_n)$.

Recursive Polar Construction

An equivalent recursive construction of the proposed scheme is provided. Similarly to the single-user construction in (5.4), the first step of the recursive construction is the composition of the wiretap channel P_2 with an input alphabet from $\mathcal{X}^2$ and an output alphabet from $\mathcal{Y}^2 \times \mathcal{Z}^2$

$$P_2(y_1, y_2, z_1, z_2|w_1, w_2) = P(y_1, z_1|w_1 + w_2)P(y_2, z_2|w_2) \tag{5.20}$$

where $(y_1, y_2) \in \mathcal{Y}^2$, $(z_1, z_2) \in \mathcal{Z}^2$, $(w_1, w_2) \in \mathcal{X}^2$, and the addition is carried modulo-2.

The continuation of the recursive construction follows in a similar manner to the recursion in Section 5.1.2; The transition probability function P_n for a channel with an input alphabet $\mathcal{X}^n$ and an output alphabet $\mathcal{Y}^n \times \mathcal{Z}^n$, is constructed using two independent copies of a channel $P_{\frac{n}{2}}$ with an input alphabet $\mathcal{X}^{\frac{n}{2}}$ and an output alphabet $\mathcal{Y}^{\frac{n}{2}} \times \mathcal{Z}^{\frac{n}{2}}$. Note that as in Section 5.1.2, all block lengths (n) are integral powers of 2. The first part of the recursive step includes the evaluation of the vectors $\mathbf{s}, \mathbf{v} \in \mathcal{X}^n$. This part is identical to the construction as described in Section 5.1.2 (steps 1 and 2). Finally, the transition probability function $P_n(\mathbf{y}|\mathbf{x})$ is given by

$$\begin{aligned} P_n(\mathbf{y}, \mathbf{z}|\mathbf{x}) =& P_{\frac{n}{2}}\left((y_1, y_2, \ldots, y_{\frac{n}{2}}), (z_1, z_2, \ldots, z_{\frac{n}{2}})|(v_1, v_2, \ldots, v_{\frac{n}{2}})\right) \\ & \cdot P_{\frac{n}{2}}\left((y_{\frac{n}{2}+1}, y_{\frac{n}{2}+2}, \ldots, y_n), (z_{\frac{n}{2}+1}, z_{\frac{n}{2}+2}, \ldots, z_n)|(v_{\frac{n}{2}+1}, v_{\frac{n}{2}+2}, \ldots, v_n)\right). \end{aligned} \tag{5.21}$$

The channel P_n in (5.21) is the combined wire-tap channel.

As in the case of standard polar coding for the single-user model, the recursive construction can be shown to be equivalent to a linear encoding with the polar generator matrix G_n:

Proposition 5.4 Let P be a binary memoryless wire-tap channel with an input alphabet $\mathcal{X}$ and output alphabets $\mathcal{Y}$ and $\mathcal{Z}$, for the legitimate user and the eavesdropper, respectively. In addition, let P_n and G_n be the combined wire-tap channel in (5.21) and the polar generator matrix in (5.6), respectively. Then,

$$P_n(\mathbf{y}, \mathbf{z}|\mathbf{w}) = P(\mathbf{y}, \mathbf{z}|\mathbf{w}G_n) \tag{5.22}$$

for all $\mathbf{w} \in \mathcal{X}^n$, $\mathbf{y} \in \mathcal{Y}^n$, and $\mathbf{z} \in \mathcal{Z}^n$.

Proof: The proof of (5.22) is identical to the proof of (5.7) in [4], where symbols from the output alphabet of the single user channel are replaced with the corresponding pair of symbols from the composite output alphabet (of the legitimate and the eavesdropper channels). ∎

To obtain the equivalence of the recursive construction of the combined channel P_n in (5.21) with the encoding operation in (5.19), the division of the components of $\mathbf{w}$ in (5.22) for information bits, random bits and predetermined and fixed bits, is detailed. This division is defined by the sets $\mathcal{A}_n$ and $\mathcal{N}_n$ as follows:

1. Over the indices specified by the index set $\mathcal{A}_n$, the information bits $\mathbf{u}$ are placed.
2. The random bits $\mathbf{b}_n^*$ are placed in the indices specified by $\mathcal{N}_n$.
3. The predetermined and fixed bits in $\mathbf{b}_n$ are left for the remaining indices specified by $\mathcal{B}_n$.

Plugging $\mathbf{u}$, $\mathbf{b}_n^*$, and $\mathbf{b}_n$ in $\mathbf{w}G_n$, results in the coded message $\mathbf{x}$ in (5.19).

Channel Splitting and Degradation Properties

The channel splitting operation in (5.10) is repeated for the case of wire-tap channels. This procedure can be carried in two different but equivalent options:

1. First performing a channel splitting operation for the wire-tap channel. This operation results in the split wire-tap channels $\{P_n^{(l)}\}_{l=1}^n$ with a binary input alphabet $\mathcal{X}$ and an output alphabet $\mathcal{Y}^n \times \mathcal{Z}^n \times \mathcal{X}^{l-1}$:

$$P_n^{(l)}(\mathbf{y}, \mathbf{z}, \mathbf{w}|w) \triangleq \frac{1}{2^{n-1}} \sum_{\mathbf{c} \in \mathcal{X}^{n-l}} P_n\big(\mathbf{y}, \mathbf{z}|(\mathbf{w}, w, \mathbf{c})\big) \tag{5.23}$$

   where $\mathbf{y} \in \mathcal{Y}^n$, $\mathbf{z} \in \mathcal{Z}^n$, $\mathbf{w} \in \mathcal{X}^{l-1}$, and $w \in \mathcal{X}$. Next, deriving the marginal split channels

$$G_n^{(l)}(\mathbf{y}, \mathbf{w}|w) \triangleq \sum_{\mathbf{z} \in \mathcal{Z}^n} P_n^{(l)}(\mathbf{y}, \mathbf{z}, \mathbf{w}|w) \tag{5.24}$$

   and

$$Q_n^{(l)}(\mathbf{z}, \mathbf{w}|w) \triangleq \sum_{\mathbf{y} \in \mathcal{Y}^n} P_n^{(l)}(\mathbf{y}, \mathbf{z}, \mathbf{w}|w) \tag{5.25}$$

   for the legitimate-user and eavesdropper, respectively, where $\mathbf{y}$, $\mathbf{z}$, $\mathbf{w}$, and w are as in (5.23).

2. First deriving the marginal combined channels:

$$G_n(\mathbf{y}|\mathbf{w}) \triangleq \sum_{\mathbf{z} \in \mathcal{Z}^n} P_n(\mathbf{y}, \mathbf{z}|\mathbf{w}) \tag{5.26}$$

   and

$$Q_n(\mathbf{z}|\mathbf{w}) \triangleq \sum_{\mathbf{y} \in \mathcal{Y}^n} P_n(\mathbf{y}, \mathbf{z}|\mathbf{w}) \tag{5.27}$$

for the legitimate user and eavesdropper, respectively, where $\mathbf{y} \in \mathcal{Y}^n$, $\mathbf{z} \in \mathcal{Z}^n$, and $\mathbf{w} \in \mathcal{X}^n$. Next, split the marginal combined channels in (5.26) and (5.27) according to

$$\frac{1}{2^{n-1}} \sum_{\mathbf{c} \in \mathcal{X}^{n-l}} G_n\big(\mathbf{y}|(\mathbf{w}, w, \mathbf{c})\big). \tag{5.28}$$

and

$$\frac{1}{2^{n-1}} \sum_{\mathbf{c} \in \mathcal{X}^{n-l}} Q_n\big(\mathbf{z}|(\mathbf{w}, w, \mathbf{c})\big) \tag{5.29}$$

where $\mathbf{y}$, $\mathbf{z}$, $\mathbf{w}$, and w are as in (5.23).

It is an immediate consequence of the equivalence properties in (5.7) and (5.22), that the split channels in (5.24) and (5.25) equal to the channels in (5.28) and (5.29).

The following proposition considers physically degraded wire-tap channels:

Proposition 5.5 Assume that the wire-tap channel P is physically degraded. Then, the split channel $P_n^{(l)}(\mathbf{y}, \mathbf{z}, \mathbf{w}|x)$ in (5.23) satisfies

$$P_n^{(l)}(\mathbf{y}, \mathbf{z}, \mathbf{w}|x) = G_n^{(l)}(\mathbf{y}, \mathbf{w}|x) D(\mathbf{z}|\mathbf{y}) \tag{5.30}$$

where $G_n^{(l)}$ is the marginal split channel of the legitimate user in (5.24), $\mathbf{y} = (y_1, \ldots, y_n) \in \mathcal{Y}^n$, $\mathbf{z} = (z_1, \ldots, z_n) \in \mathcal{Z}^n$, $\mathbf{u} \in \mathcal{X}^{l-1}$, $x \in \mathcal{X}$, $D(\mathbf{z}|\mathbf{y})$ is a memoryless transition probability law:

$$D(\mathbf{z}|\mathbf{y}) = \prod_{l=1}^{n} D(z_i|y_i)$$

and $D(z|y)$ is the conditional probability law of receiving a symbol $z \in \mathcal{Z}$ at the eavesdropper, assuming that the symbol $y \in \mathcal{Y}$ is received at the legitimate receiver.

Proof: The recursion operation in Proposition 5.2 is valid for the wire-tap channel. Specifically, for all $i > 0$ and $1 \leq l \leq 2^i$ it follows that

$$\begin{aligned} &P_{2^{i+1}}^{(2l-1)}\big((\mathbf{y}^{(1)}, \mathbf{y}^{(2)}), (\mathbf{z}^{(1)}, \mathbf{z}^{(2)}), \mathbf{w}|w_1\big) = \\ &\quad \sum_{w \in \mathcal{X}} \frac{1}{2} P_{2^i}^{(l)}\big(\mathbf{y}^{(1)}, \mathbf{z}^{(1)}, g(\mathbf{w})|w_1 + w\big) P_{2^i}^{(l)}\big(\mathbf{y}^{(2)}, \mathbf{z}^{(2)}, e(\mathbf{w})|w\big) \end{aligned} \tag{5.31}$$

$$\begin{aligned} &P_{2^{i+1}}^{(2l)}\big((\mathbf{y}^{(1)}, \mathbf{y}^{(2)}), (\mathbf{z}^{(1)}, \mathbf{z}^{(2)}), (\mathbf{w}, w_1)|w_2\big) = \\ &\quad \frac{1}{2} P_{2^i}^{(l)}\big(\mathbf{y}^{(1)}, \mathbf{z}^{(1)}, g(\mathbf{w})|w_1 + w_2\big) P_{2^i}^{(l)}\big(\mathbf{y}^{(2)}, \mathbf{z}^{(2)}, e(\mathbf{w})|w_2\big) \end{aligned} \tag{5.32}$$

where $\mathbf{y}^{(1)}, \mathbf{y}^{(2)} \in \mathcal{Y}^{2^i}$, $\mathbf{z}^{(1)}, \mathbf{z}^{(2)} \in \mathcal{Z}^{2^i}$, $\mathbf{w} \in \mathcal{X}^{2l-2}$, $w_1, w_2 \in \mathcal{X}$, and $g(\mathbf{w})$ and $e(\mathbf{w})$ are as defined in (5.14) and (5.15), respectively. The proof of the recursion property in (5.31) and (5.32) follows the exact derivation as in [4] (while replacing the output

alphabet of the single-user channel with the combined outputs of the legitimate user and the eavesdropper).

From (5.26), (5.31), and (5.32), a similar recursion follows for the marginal split channel $G_n^{(l)}(\mathbf{y}, \mathbf{w}|x)$ of the legitimate user. To this end, the recursion operations in (5.12) and (5.13) are satisfied where $p_{2^{i+1}}^{(2l-1)}$, $p_{2^i}^{(l)}$ and $p_{2^{i+1}}^{(2l)}$ are replaced by $G_{2^{i+1}}^{(2l-1)}$, $G_{2^i}^{(l)}$ and $G_{2^{i+1}}^{(2l)}$, respectively.

The proof of the degradation in (5.30) is accomplished by induction. At the first step, from (5.31) and (5.32) it follows that

$$P_2^{(1)}\big((y_1, y_2), (z_1, z_2)|w_1\big) = \sum_{w \in \mathcal{X}} \frac{1}{2} P\big(y_1, z_1|w_1 + w\big) P\big(y_2, z_2|w\big) \tag{5.33}$$

$$P_2^{(2)}\big((y_1, y_2), (z_1, z_2), w_1|w_2\big) = \frac{1}{2} P\big(y_1, z_1|w_1 + w_2\big) P\big(y_2, z_2|w_2\big). \tag{5.34}$$

Then, plugging (5.2) in (5.33) and (5.34) concludes the proof for the first step. Next, assume that the split channel $P_{2^i}^{(l)}$ satisfies the degradation property in (5.30). That is, assume that

$$P_{2^i}^{(l)}(\mathbf{y}, \mathbf{z}, \mathbf{w}'|w) = G_{2^i}^{(l)}(\mathbf{y}, \mathbf{w}'|w) D(\mathbf{z}|\mathbf{y}) \tag{5.35}$$

for all $1 \leq l \leq 2^i$, $\mathbf{y} \in \mathcal{Y}^{2^i}$, $\mathbf{z} \in \mathcal{Z}^{2^i}$, $\mathbf{w}' \in \mathcal{X}^{l-1}$, and $w \in \mathcal{X}$. Then, from (5.31) and (5.35) it follows that

$$\begin{aligned} P_{2^{i+1}}^{(2l-1)}\big((\mathbf{y}^{(1)}, \mathbf{y}^{(2)}), (\mathbf{z}^{(1)}, \mathbf{z}^{(2)}), \mathbf{w}|w_1\big) = &\sum_{w \in \mathcal{X}} \frac{1}{2} G_{2^i}^{(l)}\big(\mathbf{y}^{(1)}, g(\mathbf{w})|w_1 + w\big) D(\mathbf{z}^{(1)}|\mathbf{y}^{(1)}) \\ &G_{2^i}^{(l)}\big(\mathbf{y}^{(2)}, e(\mathbf{w})|w\big) D(\mathbf{z}^{(2)}|\mathbf{y}^{(2)}) \\ = &G_{2^{i+1}}^{(2l-1)}\big((\mathbf{y}^{(1)}, \mathbf{y}^{(2)}), \mathbf{w}|w_1\big) \\ &D\big((\mathbf{z}^{(1)}, \mathbf{z}^{(2)})|(\mathbf{y}^{(1)}, \mathbf{y}^{(2)})\big) \end{aligned}$$

where the last step follows using the recursion properties of the marginal split channel for the legitimate user. A similar argument assures the degradation property for $P_{2^{i+1}}^{(2l)}$ which concludes the proof of the proposition. ■

Successive Cancellation Decoding

The successive cancellation decoding procedure in [4] is applied for the legitimate user. The difference from the standard single-user case is that for the wire-tap channel model the legitimate user needs to decode both the message $\mathbf{u} \in \mathcal{X}^k$ and the noisy vector $\mathbf{b}_n^* \in \mathcal{X}^{k^*}$. In terms of information sets, the legitimate receiver operates on the indices specified by both $\mathcal{A}_n$ and $\mathcal{N}_n$. Denote by $\mathbf{w} = (w_1, \ldots, w_n) \in \mathcal{X}^n$ the transmitted vector over the combined channel P_n, then $\mathbf{w}$ is composed of the information vector $\mathbf{u}$, the random vector $\mathbf{b}_n^*$, and the predetermined fixed vector $\mathbf{b}_n$. It

is important not to confuse $\mathbf{w}$ with the actual codeword $\mathbf{x}$ in (5.19), which is transmitted over the given wire-tap channel P. Both interpretations are equivalent as the coset block code is equivalent to the recursive combining construction. Nevertheless, the decoding rule (and its performance analysis in the following) is characterized in terms of the vector $\mathbf{w}$, transmitted over the combined wire-tap channel and received over the marginal split channels for the legitimate user.

The decoding rule operates recursively to compute the length-n decoded vector $\hat{\mathbf{w}} = (\hat{w}_1, \ldots, \hat{w}_n) \in \mathcal{X}^n$. Let $1 \leq l \leq n$, and assume that the first $l-1$ components of $\hat{w}$, denoted by $\hat{\mathbf{w}}^{(l-1)}$, are already evaluated. If $l \notin \bar{\mathcal{A}}_n$, where

$$\bar{\mathcal{A}}_n \triangleq \mathcal{A}_n \cup \mathcal{N}_n.$$

then the current index l is not in the information index set $\mathcal{A}_n$ and not in the indices specified in $\mathcal{N}_n$ for the noisy vector. Consequently, $l \in \mathcal{B}_n$. Recall that for the indices specified by $\mathcal{B}_n$, the predetermined vector $\mathbf{b}_n$ is set. Since $\mathbf{b}_n$ is predetermined and known (both to the legitimate user and the eavesdropper), w_l is known at the receiver and therefore it is possible to set

$$\hat{w}_l = w_l.$$

If $l \in \bar{\mathcal{A}}_n$, then the current index is identified either as an information bit in $\mathbf{u}$ or as a noisy bit in $\mathbf{b}_n^*$. For this case, the following decoding rule is applied to the marginal split channel $G_n^{(l)}$ in (5.24):

$$\hat{w}_l = \begin{cases} 0 & \text{if } G_n^{(l)}(\mathbf{y}, \hat{\mathbf{w}}^{(l-1)}|0) \geq G_n^{(l)}(\mathbf{y}, \hat{\mathbf{w}}^{(l-1)}|1) \\ 1 & \text{else} \end{cases} . \tag{5.36}$$

The successive cancellation decoding described in this section, is by no means optimal. This important observation is already noted for the single-user case in [4]. Nevertheless, for an uncoded communication model with a communication channel whose transition probability function is $G_n^{(l)}$, the detection rule for the single bit w_l in (5.36) is optimal, if w_l is an equiprobable bit.

5.2.2 A Secrecy Achieving Property for Degraded Channels

Theorem 5.4 Let P be a binary-input, memoryless, degraded and symmetric wire-tap channel with a secrecy capacity $C_{\mathrm{s}}(P)$. Fix an arbitrary positive $\beta < \frac{1}{2}$, and $R < C_{\mathrm{s}}(P)$. Then, there exist sequences of sets $\mathcal{A}_n$ and $\mathcal{N}_n$ such that the secrecy coding scheme $\mathcal{C}_n(\mathcal{A}_n, \mathcal{N}_n)$ satisfies the following properties:

1. Rate: For a sufficiently large block length n

$$R \leq \frac{1}{n}|\mathcal{A}_n|. \tag{5.37}$$

2. Security: The equivocation rate $R_{\mathrm{e}}(\mathcal{C}_n(\mathcal{A}_n,\mathcal{N}_n)$ satisfies

$$\lim_{n\to\infty} R_{\mathrm{e}}\big(\mathcal{C}_n(\mathcal{A}_n,\mathcal{N}_n)\big) \geq R. \tag{5.38}$$

3. Reliability: The average block error probability under successive cancellation decoding $P_{\mathrm{e}}(\mathcal{C}_n(\mathcal{A},\mathcal{N}))$ satisfies

$$P_{\mathrm{e}}\big(\mathcal{C}_n(\mathcal{A}_n,\mathcal{N}_n)\big) = o\left(2^{-n^\beta}\right).$$

Proof:

The proof comprises of three parts: A code construction part where the construction of the sets $\mathcal{A}_n$ and $\mathcal{N}_n$ is described in detail, along with the derivation of the coding rate property in (5.37). An analysis of the equivocation rate is provided in the second part of the proof. Finally, in the third part an upper bound on the block error probability at the legitimate receiver is provided under successive cancellation decoding.

Part I: The code construction

Fix some $r^* = C(P_{Z|X}) - \epsilon$, and $r = C(P_{Y|X}) - \epsilon$, where $C(P_{Y|X})$ and $C(P_{Z|X})$ are the channel capacities of the marginal channels for the legitimate user and the eavesdropper, and $\epsilon > 0$ is determined later. According to Theorem 5.3, there exists a sequence of index sets $\tilde{\mathcal{N}}_n \subset [n]$, satisfying:

1. The cardinality of the index set $\tilde{\mathcal{N}}_n$ satisfies

$$|\tilde{\mathcal{N}}_n| \geq \lfloor nr^* \rfloor. \tag{5.39}$$

2. For all $l \in \tilde{\mathcal{N}}$, the Bhattacharyya parameter $B(Q_n^{(l)})$ of the split channel $Q_n^{(l)}$ of the eavesdropper in (5.25), is upper bounded by

$$B(Q_n^{(l)}) \leq 2^{-n^\beta}. \tag{5.40}$$

The index set $\mathcal{N}_n$ of size $\lfloor nr^* \rfloor$ is chosen arbitrary from $\tilde{\mathcal{N}}_n$.

Next, Theorem 5.3 is applied for the marginal channel of the legitimate user. Accordingly, there exists a sequence of index sets $\tilde{\mathcal{A}}_n \subset [n]$, satisfying:

1. The cardinality of the index set $\tilde{\mathcal{A}}_n$ satisfies

$$|\tilde{\mathcal{A}}_n| \geq \lfloor nr \rfloor. \tag{5.41}$$

2. For all $l \in \tilde{\mathcal{A}}_n$, the Bhattacharyya parameter $B(G_n^{(l)})$ of the split channel $G_n^{(l)}$ of the legitimate user in (5.28), is upper bounded according to

$$B(G_n^{(l)}) \leq 2^{-n^\beta}. \tag{5.42}$$

For each n, the information index set $\mathcal{A}_n$ of size $\lfloor nr \rfloor - \lfloor nr^* \rfloor$ is chosen from $\tilde{\mathcal{A}}_n \setminus \mathcal{N}_n$. As $|\mathcal{N}_n| = \lfloor nr^* \rfloor$ and $|\tilde{\mathcal{A}}_n| \geq \lfloor nr \rfloor$, the set $\tilde{\mathcal{A}}_n \setminus \mathcal{N}_n$ is of sufficient size. The specific choice of $\mathcal{A}_n$ may be carried arbitrarily. Nevertheless, the best choice is to pick the indices in $\tilde{\mathcal{A}}_n \setminus \mathcal{N}_n$ whose corresponding marginal split-channels for the legitimate-user have the lowest Bhattacharyya parameters.

The code rate of the resulting scheme satisfies

$$\begin{aligned}\frac{1}{n}|\mathcal{A}_n| &\geq \frac{r-1}{n} - \frac{r^*-1}{n} \\ &= C(P_{Y|X}) - C(P_{Z|X}) - 2\epsilon - \frac{2}{n} \\ &= C_{\mathrm{s}}(P) - 2\epsilon - \frac{2}{n}\end{aligned} \tag{5.43}$$

where the last equality follows from Theorem 5.2. Consequently, for a large enough block length and a properly chosen (small) ϵ, the code rate of the proposed scheme satisfies (5.37).

The choice of the vector $\mathbf{b}_n \in \mathcal{X}^{n-k-k^*}$ may be carried arbitrarily.

Part II: The equivocation rate analysis

The confidential message vector, the transmitted codeword, and the received vector at the eavesdropper are denoted by the random vectors $\mathbf{U}$, $\mathbf{X}$, and $\mathbf{Z}$, respectively. The equivocation rate of the proposed scheme $R_{\mathrm{e}}\big(\mathcal{C}_n(\mathcal{A},\mathcal{N})\big)$ is given by

$$\begin{aligned}R_{\mathrm{e}}\big(\mathcal{C}_n(\mathcal{A},\mathcal{N})\big) =& \frac{1}{n}H(\mathbf{U}|\mathbf{Z}) \\ &= \frac{1}{n}H(\mathbf{U}) - \frac{1}{n}I(\mathbf{U};\mathbf{Z}) \\ &= \frac{1}{n}|\mathcal{A}_n| - \frac{1}{n}I(\mathbf{U};\mathbf{Z})\end{aligned} \tag{5.44}$$

Where the last equality follows since the message bit vector is of length $|\mathcal{A}_n|$ and equiprobable. Using the chain rule of mutual information

$$\begin{aligned}I(\mathbf{U},\mathbf{X};\mathbf{Z}) =& I(\mathbf{U};\mathbf{Z}) + I(\mathbf{X};\mathbf{Z}|\mathbf{U}) \\ =& I(\mathbf{X};\mathbf{Z}) + I(\mathbf{U};\mathbf{Z}|\mathbf{X}).\end{aligned}$$

Consequently,

$$\begin{aligned} I(\mathbf{U};\mathbf{Z}) =& I(\mathbf{X};\mathbf{Z}) + I(\mathbf{U};\mathbf{Z}|\mathbf{X}) - I(\mathbf{X};\mathbf{Z}|\mathbf{U}) \\ \overset{(a)}{=}& I(\mathbf{X};\mathbf{Z}) - I(\mathbf{X};\mathbf{Z}|\mathbf{U}) \\ \leq& nC(P_{Z|X}) - I(\mathbf{X};\mathbf{Z}|\mathbf{U}) \end{aligned} \tag{5.45}$$

where (a) follows since $\mathbf{U} \to \mathbf{X} \to \mathbf{Z}$ is a Markov chain which implies that $\mathbf{Z}$ and $\mathbf{U}$ are statistically independent given $\mathbf{X}$, and $C(P_{Z|X})$ is the channel capacity of the marginal channel to the eavesdropper. The conditional mutual information $I(\mathbf{X};\mathbf{Z}|\mathbf{U})$ is given by

$$\begin{aligned} I(\mathbf{X};\mathbf{Z}|\mathbf{U}) =& H(\mathbf{X}|\mathbf{U}) - H(\mathbf{X}|\mathbf{U},\mathbf{Z}) \\ \overset{(a)}{=}& |\mathcal{N}_n| - H(\mathbf{X}|\mathbf{U},\mathbf{Z}) \\ \overset{(b)}{\geq}& n\big(C(P_{Z|X}) - \epsilon\big) - 1 - H(\mathbf{X}|\mathbf{U},\mathbf{Z}) \end{aligned} \tag{5.46}$$

where (a) follows since the binary vector $\mathbf{b}^*$ is chosen uniformly at random and it is independent with the confidential message, and (b) follows since $|\mathcal{N}_n| = \lfloor nr^* \rfloor$ and $r^* = C(P_{Z|X}) - \epsilon$.

Let $P_{\mathrm{e}|\mathbf{U}}$ denote the error probability of a decoder that needs to decode $\mathbf{X}$ while having access to both the eavesdropper observation vector $\mathbf{Z}$, the confidential message vector $\mathbf{U}$, and the predetermined vector $\mathbf{b}_n$ (which is fixed, predetermined, and known to all the users in the model). Note that if both the confidential message $\mathbf{U}$ and the predetermined vector $\mathbf{b}_n$ are known at the receiver, then the remaining uncertainty in the codeword $\mathbf{X}$ relates only to the random vector $\mathbf{b}_n^*$ of size $\mathcal{N}_n$. Using Fano's inequality (see, e.g., [23]), the conditional entropy $H(\mathbf{X}|\mathbf{U},\mathbf{Z})$ is bounded according to

$$\begin{aligned} H(\mathbf{X}|\mathbf{U},\mathbf{Z}) \leq& h_2(P_{\mathrm{e}|\mathbf{U}}) + P_{\mathrm{e}|\mathbf{U}} \log(2^{|\mathcal{N}_n|} - 1) \\ \leq& h_2(P_{\mathrm{e}|\mathbf{U}}) + nr^* P_{\mathrm{e}|\mathbf{U}} \end{aligned} \tag{5.47}$$

where $h_2(x) \triangleq -x \log x - (1-x)\log(1-x)$ is the binary entropy function. From (5.44)-(5.47) it follows that

$$R_{\mathrm{e}}\big(\mathcal{C}_n(\mathcal{A},\mathcal{N})\big) \geq \frac{1}{n}|\mathcal{A}_n| - \epsilon - \frac{1}{n} - \frac{1}{n}\left(h_2\left(P_{\mathrm{e}|\mathbf{U}}\right) + nr^* P_{\mathrm{e}|\mathbf{U}}\right) \tag{5.48}$$

$$\geq R - \frac{1}{n} - \frac{1}{n}\left(h_2\left(P_{\mathrm{e}|\mathbf{U}}\right) + nr^* P_{\mathrm{e}|\mathbf{U}}\right) \tag{5.49}$$

where the last inequality follows from (5.43) for a sufficiently small ϵ and a sufficiently large n. The error probability $P_{\mathrm{e}|\mathbf{U}}$ in (5.49) can be upper bounded by the error

probability under the suboptimal successive cancellation decoder in [4], which is fully informed with both the predetermined vector $\mathbf{b}_n$ and the confidential message vector $\mathbf{U}$. It follows from [6] that

$$P_{\mathrm{e}|\mathbf{U}} \leq o(2^{-n^\beta})$$

which concludes the proof of (5.38).

Part III: The error performance at the legitimate decoder

The successive cancellation decoding procedure at the legitimate receiver is analyzed. First, fix a vector $\mathbf{w} = (w_1, \ldots, w_n) \in \mathcal{X}^n$ comprises of the information message $\mathbf{u} \in \mathcal{X}^k$, the randomly chosen vector $\mathbf{b}^* \in \mathcal{X}^{k^*}$, and the predetermined vector $\mathbf{b} \in \mathcal{X}^{n-k-k^*}$. The conditional block error probability is denoted by $P_{\mathrm{e}|\mathbf{w}}$. That is, $P_{\mathrm{e}|\mathbf{w}}$ is the probability of a block error event given that the input vector is $\mathbf{w}$. Denote by $\mathbf{w}^{(l)} = (w_1, \ldots, w_l)$ the first l bits of $\mathbf{w}$, and by $\hat{\mathbf{w}}^{(l)} = (\hat{w}_1, \ldots, \hat{w}_l)$ the first l decoded bits. The event

$$\mathcal{F}_l \triangleq \left\{ \mathbf{w}^{(l-1)} = \hat{\mathbf{w}}^{(l-1)}, \quad w_l \neq \hat{w}_l \right\}$$

corresponds to the case where the first $l-1$ bits of $\mathbf{w}$ are decoded correctly and the first decoding error is in the l-th bit. Notice that

$$\mathcal{F}_l \subset \mathcal{E}_l(G_n^l)$$

where $\mathcal{E}_l$ is the event defined in (5.16), and G_n^l is the marginal split channel in (5.24). Consequently, it follows using the union bound that

$$\begin{aligned} P_{\mathrm{e}|\mathbf{w}} &= \Pr\left(\cup_{l=1}^n \mathcal{F}_l |\ \mathbf{w}\right) \\ &\leq \sum_{l \in \bar{\mathcal{A}}_n} \Pr\left(\mathcal{E}_l(G_n^{(l)}) |\ \mathbf{w}\right). \end{aligned} \tag{5.50}$$

Next, the summation in (5.50) is split to two summations: a summation over the indices in $\mathcal{A}_n$ and a summation over the indices in $\mathcal{N}_n$. For an index $l \in \mathcal{A}_n$, it follows from Proposition 5.3 that for all $\mathbf{w} \in \mathcal{X}^n$

$$\Pr\left(\mathcal{E}_l(G_n^{(l)}) |\ \mathbf{w}\right) \leq B(G_n^{(l)}) \tag{5.51}$$

where $B(G_n^{(l)})$ is the Bhattacharyya parameter in (5.11). To address the probability of the event $\mathcal{E}_l(G_n^{(l)})$ where $l \in \mathcal{N}_n$, notice that at the output of the marginal split channel, the decoding rule for w_l in (5.36) is optimal[1]. Recall the degradation property in Proposition 5.5. According to Proposition 5.5 the marginal split channel of

[1] As stated, this optimality is only under the setting of the split channel, and by no means implies optimality of the complete procedure (which is clearly suboptimal).

the eavesdropper is physically degraded with respect to the marginal split channel of the legitimate user. Consequently, it is clearly suboptimal to first degrade the observations at the split channel of the legitimate user, and only then to detect the bit w_l over the corresponding marginal split channel of the eavesdropper. Specifically, w_l is detected according to

$$\hat{w}_l = \begin{cases} 0 & \text{if } Q_n^{(l)}(\mathbf{z}, \hat{w}^{(l-1)}|0) \geq Q_n^{(l)}(\mathbf{z}, \hat{w}^{(l-1)}|1) \\ 1 & \text{else} \end{cases}$$

where $\mathbf{z} \in \mathcal{Z}^n$ is a degraded version of $\mathbf{y} \in \mathcal{Y}^n$, randomly picked according to the probability law $D(\mathbf{z}|\mathbf{y})$ in (5.30). This detection rule is inferior with respect to (5.36). Hence, based on Proposition 5.3, the upper bound

$$\Pr\big(\mathcal{E}_l(G_n^{(l)})|\ \mathbf{w}\big) \leq B(Q_n^{(l)}) \tag{5.52}$$

holds for all $l \in \mathcal{N}_n$. From (5.50), (5.51), and (5.52), it follows that the average block error probability is upper bounded by

$$P_{\mathrm{e}}(\mathcal{C}_n(\mathcal{A}, \mathcal{N})) \leq \sum_{l \in \mathcal{A}_n} B(G_n^{(l)}) + \sum_{l \in \mathcal{N}_n} B(Q_n^{(l)}).$$

The proof concludes using the bound on the polarization rate of the Bhattacharyya parameter in Theorem 5.3 and the specific choice of the sets $\mathcal{A}_n$ and $\mathcal{N}_n$. ■

Remark 5.3 (On communicating with full capacity) The noisy bits $\mathbf{b}_n^*$, defining the coset block code $\mathcal{C}_n^*$ based on the noisy index set $\mathcal{N}_n$ (see eq. (5.18)), are reliably detected by the legitimate user. It is therefore suggested to utilize these bits in order to communicate with the legitimate user. That is, instead of setting the bits in $\mathbf{b}_n^*$ to noisy random bits, non-secret information bits are suggested to be set on $\mathbf{b}_n^*$. The non-secret information bits must be statistically independent and equiprobable. In addition, the non-secret information must be statistically independent with the secret-information. These statistical properties allows the non-secret information bits to act as if they are noisy bits (where the eavesdropper is concerned). As a result of the cardinality of the index set $\mathcal{A}_n$ (5.41), the overall rate, including secret and non-secret information, is arbitrarily close the full (marginal) channel capacity of the legitimate user $C(P_{Y|X})$.

Remark 5.4 (The noisy bits must not be fixed) It is important to note that the bits in $\mathbf{b}_n^*$ must be chosen at random for each block transmission. To see this, first note (based on the data processing inequality) that

$$\frac{1}{n} I(\mathbf{b}_n^*; \mathbf{Z}) \leq \frac{1}{n} I(\mathbf{X}; \mathbf{Z}) \tag{5.53}$$

for all $n > 0$. Assuming that (5.53) is satisfied with equality. It follows that both the legitimate user and the eavesdropper can reliably decoded the vector $\mathbf{b}_n^*$. Considering the current setting as if it is a broadcast communication problem over the given channel, a broadcast scheme is therefore provided where we can reliably communicated with the legitimate user at a rate arbitrarily close to its marginal capacity $C(P_{Y|X})$ and at the same time with the eavesdropper at a (common) rate which is arbitrarily close to $\frac{1}{n}I(\mathbf{X};\mathbf{Z})$. This violates the fundamental limit imposed by the capacity region of the degraded broadcast channel (see, e.g., [23]). Consequently, it follows that

$$\frac{1}{n}I(\mathbf{b}_n^*;\mathbf{Z}) < \frac{1}{n}I(\mathbf{X};\mathbf{Z}) \tag{5.54}$$

for all $n > 0$. Next, since there is a one-to-one correspondence between the transmitted codeword $\mathbf{X}$ and the vector pair which is comprised of the random bits $\mathbf{b}^*$ and the confidential message $\mathbf{U}$ (the vector $\mathbf{b}$ is predetermined and fixed), it follows that

$$\begin{aligned}\frac{1}{n}I(\mathbf{X};\mathbf{Z}) &= \frac{1}{n}I(\mathbf{U},\mathbf{b}^*;\mathbf{Z}) \\ &\overset{\text{(a)}}{=} \frac{1}{n}I(\mathbf{b}^*;\mathbf{Z}) + \frac{1}{n}I(\mathbf{U};\mathbf{Z}|\mathbf{b}^*)\end{aligned} \tag{5.55}$$

for all $n > 0$, where (a) follows by the chain rule of mutual information. Hence it is observed from (5.54) and (5.55) that

$$\frac{1}{n}I(\mathbf{U};\mathbf{Z}|\mathbf{b}^*) > 0$$

for all n. This assures that if the vector $\mathbf{b}^*$ is known to the eavesdropper, for example by choosing a fixed $\mathbf{b}^*$, perfect secrecy can not be established, not even in the weak sense.

It is observed in [24], that if (R_1, R_1) is an achievable rate-equivocation pair and in addition, an additional information rate R_2 is achievable without secrecy (that is, in the ordinary notion of reliable communication), then the $(R_1 + R_2, R_1)$ is also an achieved rate-equivocation pair. The other direction is also provided in [24, p. 411]. Following Remark 5.3 which suggests the option of communicating in full rate, and the observations in [24], it is expected that the entire rate-equivocation region is obtained with polar coding. This result is provided in the following corollary:

Corollary 5.1 Under the assumptions and notation in Theorem 5.4, the entire rate-equivocation region is achievable with polar coding.

Proof: Take a rate-equivocation pair (R, R_{e}) in the rate-equivocation region defined in (5.3). Define $R_1 = R_{\mathrm{e}}$, and $R_2 = R - R_1$. Note that $R_2 \geq 0$ as $R_{\mathrm{e}} \leq R$. Consider

the coset block code in (5.18). Since $R_{\mathrm{e}} \leq C_s(P)$, the rate R_1 is achievable via the index set $\mathcal{A}_n$. It is further detailed in the proof of Theorem 5.4, that the information transmitted via the indices in $\mathcal{A}_n$ is secure. Specifically, it follows from (5.48) that the equivocation rate is arbitrarily close to $\frac{1}{n}|\mathcal{A}_n|$. As explained in Remark 5.3, reliable communication (not necessarily secure) of an additional rate of up to the capacity $C(P_{Y|X})$ of the marginal channel to the legitimate user, is achievable. Therefore, the additional rate R_2, is achievable either via the remaining indices in $\mathcal{A}_n$ and the vector $\mathbf{b}_n^*$ corresponding to the indices in $\mathcal{N}_n$. ■

5.2.3 Secrecy Achieving Properties for Erasure Wiretap Channels

In this section, a particular case of binary erasure wiretap channel is considered. Specifically, it is assumed that the channel to the legitimate user is noiseless, and the channel to the eavesdropper is a binary erasure channel (BEC) with an erasure probability δ, is considered. Recall that the set sequence $\mathcal{N}_n$ of the indices that correspond to "good" split channel to the eavesdropper, is chosen as to achieve the capacity to the eavesdropper. As the channel to the legitimate user is noiseless, that is $\mathbf{y} = \mathbf{x}$, the set sequence $\mathcal{A}_n$ and is set according to

$$\mathcal{A}_n \triangleq [n] \setminus \mathcal{N}_n. \tag{5.56}$$

Note that for this particular case $\mathcal{B}_n = \emptyset$. The resulting coding scheme is then a particular case of the coset coding scheme in [82] where the base code is determine by the generator matrix $G_n(\mathcal{N}_n)$ and the actual coset is determined by $\mathbf{u}G_n(\mathcal{A}_n)$ where $\mathbf{u}$ is the transmitted information bits (the secret message) and G_n is the polar generator matrix for a block length n. Specifically, the codeword $\mathbf{x}$ is given, based on(5.19), by

$$\mathbf{x} = \mathbf{u}G_n(\mathcal{A}_n) + \mathbf{b}_n^* G_n(\mathcal{N}_n). \tag{5.57}$$

The rate and reliability properties in this particular case follows immediately as a result of Theorem 5.4. That is, the rate approaches the secrecy capacity, which in this case equals δ, and the legitimate user obviously can decode the transmitted message. As in the second part of the proof of Theorem 5.4, the confidential message vector, the transmitted codeword, and the received vector at the eavesdropper are denoted by the random vectors $\mathbf{U}$, $\mathbf{X}$, and $\mathbf{Z}$, respectively. The following lemma address the entropy measure $H(\mathbf{U}|\mathbf{Z})$.

Lemma 5.1 Under the assumption and notation for the consider binary erasure wiretap channel, the entropy $H(\mathbf{U}|\mathbf{Z})$ satisfies

$$H(\mathbf{U}|\mathbf{Z}) \geq n\delta(1 - c2^{-n^{\beta}})$$

where δ is the erasure probability of the wiretap channel, and $c > 0$.

Proof: Let us fix a particular realization of the channel erasure sequence [2]. Denote by $\mathcal{D}$ the set of μ indices which are not erased. That is, the eavesdropper received the bits X_i for every $i \in \mathcal{D}$, and erasure symbols for every index in $\mathcal{D}^{\mathrm{c}} \triangleq [n] \setminus \mathcal{D}$. Consider the $|\mathcal{N}_n| \times n$ matrix $\{G_n(\mathcal{N}_n)\}$. As the generator matrix G_n for the polar construction has a full rank (for every n in the construction), the matrix $G_n(\mathcal{N}_n)$ has a rank $\mathcal{N}_n$. Therefore, it is a generator matrix for a binary linear block code of dimension $\mathcal{N}_n$. This code has a parity check matrix of size $|\mathcal{A}_n| \times n$, denoted by H_n (recall that $\mathcal{A}$ is given by (5.56)). Since all the information bits are equiprobable, and all the noisy bits are also equiprobable, the codeword $\mathbf{X}$, given by (5.19), id uniformly distributed over all possible binary vectors in $\{0,1\}^n$. Consequently, all the bits in $\mathbf{X}$ are independent and identically distributed uniform binary random variables. Hence, $H(\mathbf{X}|\mathbf{Z}) = n-\mu$. In addition, note that if the codeword $\mathbf{X}$ is known, then information bits $\mathbf{U}$ are fully determined for the considered polar coding scheme. It follows that

$$H(\mathbf{U}|\mathbf{Z}) = H(\mathbf{U}|\mathbf{X},\mathbf{Z}) + H(\mathbf{X}|\mathbf{Z}) - H(\mathbf{X}|\mathbf{U},\mathbf{Z}) \tag{5.58}$$

$$= m - \mu - H(\mathbf{X}|\mathbf{U},\mathbf{Z}). \tag{5.59}$$

Note that (5.58) is a restatement of [82, Eq. (5)], and (5.59) is a restatement of [82, Eq. (6)].

Next, fix a realization $\mathbf{Z} = \mathbf{z} \in \{0,1\}^n$ and $\mathbf{U} = \mathbf{u} \in \{0,1\}^{|\mathcal{A}_n|}$. From (5.57), it follows that the erased bits $\{X_i\}_{i\in\mathcal{D}^{\mathrm{c}}}$ satisfies the linear equations

$$\sum_{i\in\mathcal{D}} X_i \left(H_n\right)_i = H_n \mathbf{u} G_n\left(\mathcal{A}_n\right) + \sum_{i\in\mathcal{D}^{\mathrm{c}}} X_i \left(H_n\right)_i \tag{5.60}$$

where $\left(H_n\right)_i$ is the i-th column of the parity check matrix H_n. The number of solutions to (5.60) is given by

$$2^{n-\mu-d\left(\left\{(H_n)_i\right\}_{i\in\mathcal{D}}\right)}$$

where $d\left(\{(H_n)_i\}_{i\in\mathcal{D}}\right)$ is the dimension of the linear space spanned by the the column vectors in $\{(H_n)_i\}_{i\in\mathcal{D}}$. Since all the solutions for the erasures X_i, $i \in \mathcal{D}$, are equally likely, it follows that

$$H(\mathbf{X}|\mathbf{U}=\mathbf{u},\mathbf{Z}=\mathbf{z}) = n - \mu - d\left(\{(H_n)_i\}_{i\in\mathcal{D}}\right). \tag{5.61}$$

[2] This case is studied in [82], and some parts of the provided proof are based on proper presentation of the techniques developed in [82] for the case at hand.

From (5.59) and (5.61), it follows that

$$H(\mathbf{U}|\mathbf{Z}) = \mathsf{E}d\left(\{(H_n)_i\}_{i\in\mathcal{D}}\right). \tag{5.62}$$

As the information indices $\mathcal{N}_n$ for the eavesdropper are chosen such that it can decode the noisy bits $\mathbf{b}^*$ with an error probability of $\mathcal{O}(2^{-n^\beta})$, it follows that

$$H(\mathbf{U}|\mathbf{Z}) \geq \mathsf{E}\left(d\left(\{(H_n)_i\}_{i\in\mathcal{D}}\right)|, \text{correct decoding}\right)(1-c2^{-n^\beta}) \tag{5.63}$$

$$= n\delta(1-c2^{-n^\beta}) \tag{5.64}$$

where $c > 0$ and δ is the erasure probability of the eavesdropper channel. ■

Remark 5.5 (**All coset must be equally likely**) In the current discussion, the secrecy polar coding scheme is applied with $\mathcal{B}_n = \emptyset$. This fact is crucial for the proof of Lemma 5.1. It is conjectured that this choice may be crucial to achieve the entire secrecy capacity under the strong secrecy condition.

Remark 5.6 (**On possible stronger notion of secrecy**) Consider the conditions in Theorem 5.3. In particular, not that the rate $R < C(p)$ is kept fixed for the polarization structure of the code. If, it be possible to construct the sequence of polar codes, with a sequence of blocklength dependent rates R_n having the property that

$$R_n \geq C(p) - \frac{\alpha}{n^\gamma} \tag{5.65}$$

where $\alpha > 0$ and $\gamma > 1$ are arbitrarily fixed parameters. Then, it will follow as a corollary of Lemma 5.1 that a strong notion of secrecy is guaranteed. That is, the entropy $H(\mathbf{U}|\mathbf{Z})$ is arbitrarily close to $H(\mathbf{U})$. To see this, note that if polarization is possible while satisfying (5.65), it follows that

$$|\mathcal{N}_n| \geq n\left(1-\delta-\frac{\alpha}{n^\gamma}\right).$$

Consequently,

$$H(\mathbf{U}) = |\mathcal{A}_n| = n - |\mathcal{N}_n| = n\delta + \frac{\alpha}{n^{1-\gamma}}.$$

Hence $H(\mathbf{U}|\mathbf{Z})$ is lower bounded by a quantity which is arbitrarily close $H(\mathbf{U})$ as the blocklength increases. For the particular case of the BEC, it follows from [4, Eq. (34)-(35)], that the considered question requires the analysis of the following sequence

$$|\{i \in [n] : \ Z_n^i \leq Ce^{n^\beta}\}|$$

where $\{Z_n^i\}_{i\in[n]}$ is a sequence, generated recursively according to

$$Z_{2k}^{(2i-1)} = 2Z_k^{(i)} - \left(Z_k^{(i)}\right)^2$$

$$Z_{2k}^{(2i)} = \left(Z_k^{(i)}\right)^2.$$

where $i \in [k]$ and $Z_1^{(1)} = \delta$.

5.3 An open polarization problem and the general wiretap channel

An open polarization problem is presented in addition to a conjecture which suggests a possible solution. A polar secrecy scheme for non-degraded wiretap channels is provided based on suggested conjecture.

5.3.1 On the polarization of the 'bad' indices

Let $\mathbf{W} = (W_1, \ldots, W_n)$ be a random vector, where $\{W_i\}_{i=1}^n$ are statistically independent and equiprobable $\Pr(W_i = 0) = \Pr(W_i = 1) = \frac{1}{2}$ for all $i \in [n]$. The random vector $\mathbf{W}$ is polar encoded to a codeword $\mathbf{X} = G_n\mathbf{W}$, where G_n is the polar generator matrix of size n. The codeword $\mathbf{X}$ is transmitted over a binary input DMC p, whose output alphabet is $\mathcal{Y}$. The received vector is denoted by $\mathbf{Y} = (Y_1, \ldots, Y_n)$. For a given vector $\mathbf{W}$ and a set $\mathcal{A} \subseteq [n]$, the following notation is used

$$\mathbf{W}_{\mathcal{A}} \triangleq (W_{i_1}, \ldots W_{i_{|\mathcal{A}|}})$$

where $i_1 < i_2 < \ldots < i_{|\mathcal{A}|}$ and $i_k \in \mathcal{A}$ for all $k \in [|\mathcal{A}|]$. Define the following quantities of mutual information

$$I_i \triangleq I(W_i; \mathbf{W}_{[i-1]}, \mathbf{Y}), \quad i \in [n]. \tag{5.66}$$

The following polarization of mutual information is the key result in [4], [6]:

Theorem 5.5 (On the polarization of mutual information [4]) Assume that p is a binary-input output-symmetric DMC whose capacity is $C(p)$, and fix $0 < \delta < 1$. Then,

$$\lim_{n\to\infty} \left(\frac{1}{n} \Big| \big\{ i \in [n] : \ I_i \in (1-\delta, 1] \big\} \Big| \right) = C(p)$$
$$\lim_{n\to\infty} \left(\frac{1}{n} \Big| \big\{ i \in [n] : \ I_i \in [0, \delta) \big\} \Big| \right) = 1 - C(p).$$

Denote by $\mathcal{A}_n$ the set of indices for which the corresponding mutual information quantities I_i, $i \in \mathcal{A}_n$, are arbitrarily close to 1 bit (for a sufficiently large n). The set $\mathcal{A}_n$ is called the information index set. This is the very same index set in Theorem 5.3, of 'good' split channels whose corresponding Bhattacharyya constants approach 0. Let $\mathcal{A}'_n \subset \mathcal{A}_n$ and let $\mathcal{S}_n \subseteq \mathcal{A}_n^{\mathrm{c}}$. We define the index sets

$$\mathcal{D}_n \triangleq \mathcal{A}'_n \cup \mathcal{S}_n$$

and

$$\mathcal{D}_n^{(i)} \triangleq \big\{ j \in D_n : \ j < i \big\}, \quad i \in [n].$$

A problem of interest lies in the $|\mathcal{D}_n|$ quantities of mutual information:

$$J_i \triangleq I(W_i; \mathbf{W}_{\mathcal{D}_n^{(i)}}, \mathbf{W}_{\mathcal{D}_n^{\mathrm{c}}}, \mathbf{Y}), \quad i \in \mathcal{D}_n. \tag{5.67}$$

For the indices in $\mathcal{A}'_n$ a straight froward answer is provided:

Lemma 5.2 (on the indices of 'good' split channels) Fix a $0 < \delta < 1$ and an index $i \in \mathcal{A}'_n$. For sufficiently large n

$$J_i \geq 1 - \delta.$$

Proof: As the mutual information I_i in (5.66) includes a subset of the random variables in J_i in (5.67), it follows that

$$J_i \geq I_i.$$

The proof concludes using Theorem 5.5 as $\mathcal{A}'_n \subset \mathcal{A}_n$. ■

According to Lemma 5.2 'good' indices for which the mutual information quantities I_i approach 1 bit, remain 'good' with respect to the mutual information J_i. The characterization of the 'bad' indices seems at this point to be a greater challenge. A conjecture for possible polarization properties of the mutual information quantities J_i in (5.67) is provided for the ('bad') indices in $\mathcal{S}_n$. Two possible polarization properties are considered:

Conjecture 1 (On possible polarization dichotomy) Fix a $0 < \delta < 1$. There exists a partition of $\mathcal{S}_n$ to two sets $\mathcal{S}'_n$ and $\mathcal{S}''_n = \mathcal{S}_n \setminus \mathcal{S}'_n$, such that for a sufficiently large n

$$J_i < \delta, \quad \text{for all } i \in S'_n \tag{5.68}$$

$$J_i > 1 - \delta, \quad \text{for all } i \in S''_n. \tag{5.69}$$

Remark 5.7 (On degenerated and non-degenerated possible partitions) One of the possible option resulting from Conjecture 1 is that $\mathcal{S}'_n = \mathcal{S}_n$. In case where this degenerated partition is proved to be correct, then it follows that the additional information provided by the bits in $\mathbf{W}_{\mathcal{D}_n^{\mathrm{c}}}$ do not alter the known polarization of the mutual information quantities I_i in (5.66). The non-degenerated partition of $\mathcal{S}_n$ offers (in the case it is proven to be correct) a dichotomy of the indices in $\mathcal{S}_n$. Accordingly, either the former polarization remains or alternatively the knowledge of the bits in

$\mathbf{W}_{\mathcal{D}_n^c}$ completely changes the orientation of the polarization. The size of $\mathcal{S}_n'' \cup \mathcal{A}_n'$ must satisfy

$$|\mathcal{S}_n'' \cup \mathcal{A}_n'| \overset{(a)}{=} |\mathcal{A}_n'| + |\mathcal{A}_n''| \overset{(b)}{\leq} nC(p). \tag{5.70}$$

Equality (a) in (5.70) is obvious as the sets $\mathcal{A}_n'$ and $\mathcal{S}_n$ are disjoint. Violating the inequality (b) in (5.70) results in violating the coding theorem for a DMC as the input bits to the split channels specified by the set $\mathcal{S}_n'' \cup \mathcal{A}_n'$ can be reliably decoded (This can be shown in a similar fashion as in [4]).

Remark 5.8 (On a particular trivial case where Conjecture 1 is true) There exists an option where Conjecture1 is trivially proved as a particular application of Theorem 5.5. Specifically, assume that for every index $i \in \mathcal{D}_n$, it follows that

$$j < i \quad \forall j \in \mathcal{D}_n^c.$$

In that case, the degenerated partition in Remark 5.7 follows as an immediate particular case of Theorem 5.5.

5.3.2 A polar secrecy scheme

In this section, a polar secrecy scheme is provided assuming that Conjecture 1 is true. The same notation and definitions of the coset code defined in Section 5.2.1 are assumed. The transmitted codeword $\mathbf{x}$ is defined in (5.19). This definition is based on the index sets $\mathcal{A}_n$ and $\mathcal{N}_n$. The secure information bits are considered as if they are being transmitted over the split channels whose indices are in $\mathcal{A}_n$. Over the split channels whose indices are in $\mathcal{N}_n$, noisy bits are attributed. The polar secrecy scheme is provided in Section 6.3 by a proper choice of the sets $\mathcal{A}_n$ and $\mathcal{N}_n$. The degradation property in Section 6.3 assures that the indices which correspond to split channels which polarize to 'good channels' for the eavesdropper, also polarize for 'good channels' for the legitimate user. This clearly does not necessarily follow for the general not-degraded case.

For the general wiretap channel, indices that are 'good' for the eavesdropper may not be 'good' for the legitimate user and vice-versa. A binary-input symmetric wiretap channel is assumed. As in the construction detailed in Part I of the proof of Theorem 5.4, the sets $\tilde{\mathcal{A}}_n$ and $\tilde{\mathcal{N}}_n$ of 'good indices' are considered. The sets $\tilde{\mathcal{A}}_n$ and $\tilde{\mathcal{N}}_n$ include the indices for which the Bhattacharyya parameters of the corresponding split channels approach zero as the block length approach infinity. Specifically, fixing $r < C(P_{Y|X})$ and $r^* < C(P_{Z|X})$, the conditions in (5.39)-(5.42) follow.

Define the index set $\mathcal{S}_n \triangleq \tilde{\mathcal{A}}_n \setminus \tilde{\mathcal{N}}_n$ of indices which are 'good' for both the legitimate user and the eavesdropper. According to Conjecture 1, the set $\mathcal{S}_n$ can

be partitioned into two index sets $\mathcal{S}'_n$ and $\mathcal{S}''_n$, satisfying the polarization properties in (5.68)-(5.69) where $\mathcal{A}_n$ is replaced by $\tilde{\mathcal{N}}_n$, and $\mathcal{A}'_n$ is replaced by $\tilde{\mathcal{A}}_n \cap \tilde{\mathcal{N}}_n$. Next, the set $\mathcal{N}_n$ is defined according to

$$\mathcal{N}_n \triangleq \left(\tilde{\mathcal{A}}_n \cap \tilde{\mathcal{N}}_n\right) \cup \mathcal{S}''_n \tag{5.71}$$

and the set $\mathcal{A}_n$ is defined to be the remaining indices in $\mathcal{S}_n$, that is

$$\mathcal{A}_n \triangleq \mathcal{S}'_n.$$

As explained in Remark 5.7, the term $\frac{1}{n}|\mathcal{N}_n|$ can not exceed the capacity of the eavesdropper marginal channel. Consequently, the size of $\mathcal{S}'_n$ can be chosen such that $\frac{1}{n}|\mathcal{S}'_n|$ is arbitrarily close to $C(P_{Y|X}) - C(P_{Z|X})$.

Next, the same coset coding scheme defined in (5.19) is applied to the case at hand (with the new construction of the sets $\mathcal{A}_n$ and $\mathcal{N}_n$). As the information rate $\frac{1}{n}|\mathcal{A}_n|$ of the considered scheme may be chosen arbitrarily close to $C(P_{Y|X}) - C(P_{Z|X})$, the same coding rate as in Theorem 5.4 is obtained. The decoding reliability at the legitimate user is clear and follows the same proof as for the degraded case (note that all the noisy bits in the considered scheme are 'transmitted' over the split channels that are 'good' for the legitimate user). It is left to establish that the equivocation rate can approach the information rate of the considered scheme.

5.3.3 Analysis of the equivocation rate

As explained in Section 5.2.1, the bits $\mathbf{b}_n$ corresponding to the indices in $\mathcal{B}_n$ are predetermined and fixed. These bits are known both to the eavesdropper and the legitimate user. For each blocklength n, consider the ensemble of coset codes corresponding for all the possible selection of fixed bits $\mathbf{b}_n$. An analysis of the equivocation rate where the coset code is chosen in random is considered. Specifically, it is assumed that the actual code is chosen from the ensemble by picking the bits in $\mathbf{b}_n$ in random. The random selection of the bits in $\mathbf{b}_n$ is carried independently and identically. Each bit is picked at random with an equiprobable probability, $\Pr(0) = \Pr(1) = \frac{1}{2}$. In addition, it is assumed that the random selection of $\mathbf{b}_n$ is independent with the random noisy bits in $\mathbf{b}^*_n$ and the secret message. It is important to distinguish between the ransom selection of a code and the noisy bits $\mathbf{b}^*$. The random selection of code is part of our analysis, this selection (i.e., the bits in $\mathbf{b}_n$) is known to both the legitimate and the eavesdropper. In contrast, the random noisy bits $\mathbf{b}^*$ are immanent part of the encoding procedure and they are unknown to both the legitimate user and the receiver. The noisy bits $\mathbf{b}^*$ are picked randomly, each independent with the others, and with

a uniform probability. The information bits are also assumed to be independent and equiprobable.

The secrecy properties of the suggested scheme is considered in the following proposition:

Proposition 5.6 Consider the polar secrecy scheme in Section 5.3.2 whose transmissions take place over a binary-input memoryless symmetric wiretap channel. Then, there exists a bit vector $\mathbf{b}_n$ for which the equivocation rate satisfy the secrecy condition in (5.38).

Proof: Denote by $\mathbf{W}$ the random binary vector comprises the random bits in $\mathbf{b}_n$, $\mathbf{b}_n^*$, and $\mathbf{u}$ in the encoding procedure (5.19), and by $\mathbf{Z}$ the random vector received at the eavesdropper. According to the considered assumptions, all the bits in $\mathbf{W}$ are independent and equiprobable. It follows using the chain rule of mutual information that

$$\begin{aligned} I(\mathbf{W}_{\mathcal{N}_n}, \mathbf{W}_{\mathcal{A}_n}; \mathbf{W}_{\mathcal{B}_n}, \mathbf{Z}) &= I(\mathbf{W}_{\mathcal{A}_n}; \mathbf{W}_{\mathcal{B}_n}, \mathbf{Z}) + I(\mathbf{W}_{\mathcal{N}_n}; \mathbf{W}_{\mathcal{B}_n}, \mathbf{Z} \mid \mathbf{W}_{\mathcal{A}_n}) \\ &= I(\mathbf{W}_{\mathcal{A}_n}; \mathbf{W}_{\mathcal{B}_n}) + I(\mathbf{W}_{\mathcal{A}_n}; \mathbf{Z} \mid \mathbf{W}_{\mathcal{B}_n}) \\ &\quad + I(\mathbf{W}_{\mathcal{N}_n}; \mathbf{W}_{\mathcal{B}_n} \mid \mathbf{W}_{\mathcal{A}_n}) + I(\mathbf{W}_{\mathcal{N}_n}; \mathbf{Z} \mid \mathbf{W}_{\mathcal{A}_n}, \mathbf{W}_{\mathcal{B}_n}) \\ &= H(\mathbf{W}_{\mathcal{A}_n}) - H(\mathbf{W}_{\mathcal{A}_n} \mid \mathbf{Z}, \mathbf{W}_{\mathcal{B}_n}) + I(\mathbf{W}_{\mathcal{N}_n}; \mathbf{Z} \mid \mathbf{W}_{\mathcal{A}_n}, \mathbf{W}_{\mathcal{B}_n}) \end{aligned} \tag{5.72}$$

where the last equality follows since $\mathbf{W}_{\mathcal{A}_n}$, $\mathbf{W}_{\mathcal{N}_n}$, and $\mathbf{W}_{\mathcal{B}_n}$ are independent. As the set $\mathcal{N}_n$ comprises indices of split channels which polarize to perfect channels, the bits in $\mathbf{W}_{\mathcal{N}_n}$ can be reliably decoded at the eavesdropper based on perfect knowledge of the remaining bits and the received vector (this is shown in a similar fashion to [4]). Hence, the decoding error probability $P_{\mathrm{e}}(\mathbf{W}_{\mathcal{N}_n^{\mathrm{c}}})$ of the bits in $\mathbf{W}_{\mathcal{N}_n}$ based on the received vector and the remaining bits $\mathbf{W}_{\mathcal{N}_n^{\mathrm{c}}}$, can be made arbitrarily low. As a consequence of Fano's inequality it follows that

$$\begin{aligned} |\mathcal{N}_n| &\geq I(\mathbf{W}_{\mathcal{N}_n}; \mathbf{Z} \mid \mathbf{W}_{\mathcal{A}_n}, \mathbf{W}_{\mathcal{B}_n}) \\ &= H(\mathbf{W}_{\mathcal{N}_n} \mid \mathbf{W}_{\mathcal{A}_n}, \mathbf{W}_{\mathcal{B}_n}) - H(\mathbf{W}_{\mathcal{N}_n} \mid \mathbf{Z}, \mathbf{W}_{\mathcal{A}_n}, \mathbf{W}_{\mathcal{B}_n}) \\ &> H(\mathbf{W}_{\mathcal{N}_n}) - h_2\big(P_{\mathrm{e}}(\mathbf{W}_{\mathcal{N}_n^{\mathrm{c}}})\big) - |\mathcal{N}_n| P_{\mathrm{e}}(\mathbf{W}_{\mathcal{N}_n^{\mathrm{c}}}) \end{aligned} \tag{5.73}$$

where h_2 is the binary entropy function. For a sufficiently large block length n, the expected decoding error probability approaches zero. Consequently, the rate $\frac{1}{n} I(\mathbf{W}_{\mathcal{N}_n}; \mathbf{Z} \mid \mathbf{W}_{\mathcal{A}_n}, \mathbf{W}_{\mathcal{B}_n})$ can be made arbitrarily close to $\frac{1}{n}|\mathcal{N}_n|$. It follows from (5.72) and (5.73) that

$$\frac{1}{n} H(\mathbf{W}_{\mathcal{A}_n} \mid \mathbf{Z}, \mathbf{W}_{\mathcal{B}_n}) \geq \frac{|\mathcal{A}_n|}{n} + \frac{|\mathcal{N}_n|}{n} - \epsilon_n - \frac{1}{n} I(\mathbf{W}_{\mathcal{N}_n}, \mathbf{W}_{\mathcal{A}_n}; \mathbf{W}_{\mathcal{B}_n}, \mathbf{Z}) \tag{5.74}$$

where $\epsilon_n \geq 0$ and approaches zero as n grows.

Based on Conjecture 1, the mutual information $\frac{1}{n}I(\mathbf{W}_{\mathcal{N}_n}, \mathbf{W}_{\mathcal{A}_n}; \mathbf{W}_{\mathcal{B}_n}, \mathbf{Z})$ can be shown to be arbitrarily close to $\frac{1}{n}|\mathcal{N}_n|$. Using the chain rule of mutual information it follows that

$$\begin{aligned} I(\mathbf{W}_{\mathcal{N}_n}, \mathbf{W}_{\mathcal{A}_n}; \mathbf{W}_{\mathcal{B}_n}, \mathbf{Z}) &= \sum_{i\in\mathcal{N}_n} I(\mathbf{W}_i; \mathbf{W}_{\mathcal{B}_n}, \mathbf{Z} \mid \mathbf{W}_{\mathcal{N}_n^{(i)}}, \mathbf{W}_{\mathcal{A}_n^{(i)}}) \\ &\quad + \sum_{i\in\mathcal{A}_n} I(\mathbf{W}_i; \mathbf{W}_{\mathcal{B}_n}, \mathbf{Z} \mid \mathbf{W}_{\mathcal{N}_n^{(i)}}, \mathbf{W}_{\mathcal{A}_n^{(i)}}). \\ &= \sum_{i\in\mathcal{N}_n} I(\mathbf{W}_i; \mathbf{W}_{\mathcal{N}_n^{(i)}}, \mathbf{W}_{\mathcal{A}_n^{(i)}}, \mathbf{W}_{\mathcal{B}_n}, \mathbf{Z}) \\ &\quad + \sum_{i\in\mathcal{A}_n} I(\mathbf{W}_i; \mathbf{W}_{\mathcal{N}_n^{(i)}}, \mathbf{W}_{\mathcal{A}_n^{(i)}}, \mathbf{W}_{\mathcal{B}_n}, \mathbf{Z}). \end{aligned} \tag{5.75}$$

where the last equality follows as all the bits in $\mathbf{W}$ are independent. For every index $i \in \mathcal{N}_n$, it follows from Lemma 5.2 and Conjecture 1 that

$$I(W_i; \mathbf{W}_{\mathcal{N}_n^{(i)}}, \mathbf{W}_{\mathcal{A}_n^{(i)}}, \mathbf{W}_{\mathcal{B}_n}, \mathbf{Z}) > 1 - \delta. \tag{5.76}$$

In addition, for all the indices $i \in \mathcal{A}_n$ it also follows from Conjecture 1 that

$$I(W_i; \mathbf{W}_{\mathcal{N}_n^{(i)}}, \mathbf{W}_{\mathcal{A}_n^{(i)}}, \mathbf{W}_{\mathcal{B}_n}, \mathbf{Z}) < \delta. \tag{5.77}$$

From (5.75), (5.76) and (5.77) it follows that

$$\begin{aligned} \frac{1}{n}I(\mathbf{W}_{\mathcal{N}_n}, \mathbf{W}_{\mathcal{A}_n}; \mathbf{W}_{\mathcal{B}_n}, \mathbf{Z}) &\leq \frac{|\mathcal{N}_n|}{n} + \frac{\delta|\mathcal{A}_n|}{n} \\ &\leq \frac{|\mathcal{N}_n|}{n} + \delta. \end{aligned} \tag{5.78}$$

Hence, based on (5.74) and (5.78) we end up with

$$\frac{1}{n}H(\mathbf{W}_{\mathcal{A}_n} \mid \mathbf{Z}, \mathbf{W}_{\mathcal{B}_n}) \geq \frac{1}{n}|\mathcal{A}_n| - \epsilon_n - \delta.$$

As δ can be fixed arbitrarily small, and ϵ_n approaches zero, the equivocation rate can be made arbitrarily close to $\frac{1}{n}|\mathcal{A}_n|$ which assures the secrecy property of the provided scheme. ■

5.4 Summary and Conclusions

A polar secrecy scheme is provided in this chapter for the two-user, memoryless, symmetric and degraded wire-tap channel.The provided polar coding scheme is shown to achieve the entire rate-equivocation region for the considered communication model.

The analysis of non-degraded channel models is of great priority. In particular, proving Conjecture 1 is the main interest in the continuation of the research discussed in this chapter. The following generalizations are of additional possible interest:

1. Non-binary settings: In light of the recent results by Sasoglu et al. [91], a generalization to the non-binary setting may be a straight forward generalization.

2. Secrecy polar schemes for non-symmetric wiretap channels, based on the non-binary polarization provided in [91].

3. Polar coding for a broadcast channel with confidential messages. The particular case of degraded message sets over a degraded channel is first considered.

4. Strong secrecy properties: As noted, the provided scheme is shown to provide weak secrecy. It is of great interest to find out if this scheme can also provide strong secrecy.

5. Generalized polar secrecy-schemes based on the ideas in [5], [64]-[62].

6. Combing the polar scheme with the MAC approach for the wiretap channel (see, e.g., [83]).

Chapter 6

Parallel Polar-Coding

Chapter Overview

A parallel polar coding scheme is provided in this chapter for communicating over binary-input arbitrarily-permuted memoryless symmetric parallel-channels. In [4] where symmetric DMC are concerned, the predetermined bits may be chosen arbitrarily; they are fixed and do not depend on the transmitted message. For the scheme provided in this chapter, some of these bits incorporate an algebraic structure and depend on the transmitted message. Moreover, the determination of these bits is based on the structural properties of MDS codes, in a manner which relates to the rate-matching code in [116]. The chapter is based on the following paper:

E. Hof, I. Sason, and S. Shamai (Shitz), "Polar Coding for Reliable Communications over Parallel Channels," submitted to the *IEEE Trans. on Information Theory*, July 2010. This work is presented in part in the *2010 IEEE Information Theory Workshop (ITW 2010),* Dublin, Ireland, September 2010.

This chapter is structured as follows: Section 6.1 provides some preliminary material. Section 6.2 considers channel polarization for (stochastically) degraded parallel channels. The parallel polar coding scheme is introduced and analyzed in Section 6.3.

6.1 Preliminaries

For an introductory section on channel polarization coding the reader is referred to Section 5.1.2.

6.1.1 Arbitrarily Permuted Parallel Channels

We consider the communication model in Figure 6.1. A message m is transmitted over a set of S parallel memoryless channels. The notation

$$[S] \triangleq \{1, \ldots, S\}$$

is used in this paper. All channels are assumed to have a common input alphabet $\mathcal{X}$, and possibly different output alphabets $\mathcal{Y}_s$, $s \in [S]$. The transition probability function of each channel is denoted by $P_s(y_s|x)$, where $y_s \in \mathcal{Y}_s$, $s \in [S]$, and $x \in \mathcal{X}$. For the particular case depicted in Figure 6.1, the communication takes place over a set of $S = 3$ parallel channels. The encoding operation maps the message m into a set of S codewords $\{\mathbf{x}_s \in \mathcal{X}^n\}_{s=1}^S$. Each of these codewords is of length n, and it is transmitted over a different channel. The mapping of codewords to channels is done by an arbitrary permutation $\pi : [S] \to [S]$. The permutation π is fixed during the transmission of the codewords. The set of possible S channels are known at both the encoder and decoder. The encoder has no information about the chosen permutation. The decoder, on the other hand, knows the specific chosen permutation. The coding problem for this communication model is to guarantee reliable communication for all possible $(S!)$ permutations π. This problem is formulated and studied in [116].

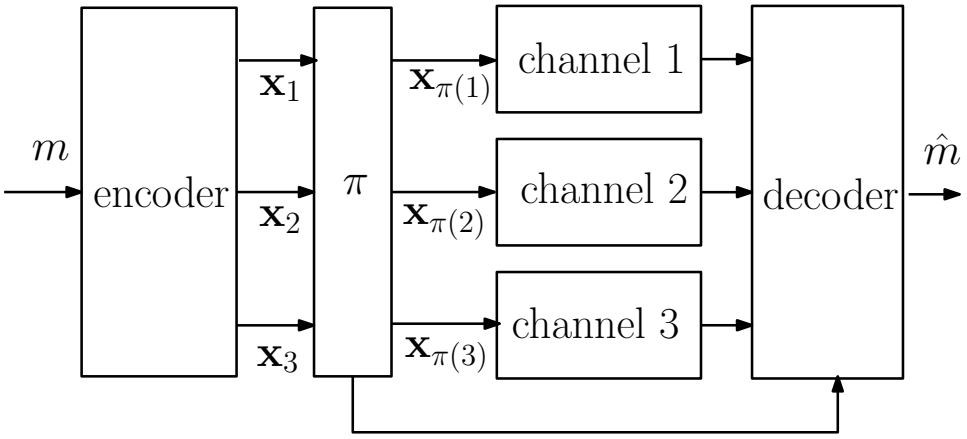


Figure 6.1: Communication over an arbitrarily-permuted parallel channel. The particular case of communicating over $S = 3$ parallel channels is depicted (taken from [116]).

Definition 6.1 (Achievable rates and channel capacity) Consider coded communication over a set of S arbitrarily permuted parallel channels. A rate $R > 0$ is achievable if there exists a sequence of encoders and decoders such that for all $\delta > 0$

and a sufficiently large block length n

$$\frac{1}{n}\log_2 M \geq R - \delta \tag{6.1}$$

$$P_{\mathrm{e}}^{(\pi)}(n) \leq \delta, \text{ for all } S! \text{ permutations } \pi : [S] \to [S] \tag{6.2}$$

where M is the number of possible messages and $P_{\mathrm{e}}^{(\pi)}(n)$ is the average block error probability for a fixed permutation π and block length n. The capacity of the considered model C_{Π} is the maximal achievable rate to satisfy (6.1) and (6.2).

Theorem 6.1 (The capacity of arbitrarily-permutated memoryless parallel channels [116]) Consider the transmission over a set of S arbitrarily-permutated memoryless parallel channels. Assume that there is an input distribution that achieves capacity for all parallel channels. Then, the capacity C_{Π} satisfies

$$C_{\Pi} = \sum_{s=1}^{S} C_s \tag{6.3}$$

where C_s is the capacity of the s-th channel, $s \in [S]$.

Remark 6.1 In case that there is no common input distribution that achieves the capacity of each component channel, see [116, Theorem 2, Eq. 50].

As noted in [116], $\sum_{s=1}^{S} C_s$ is the capacity if both the encoder and decoder know the actual permutation π; since the encoder does not know the actual permutation, then $C \leq \sum_{s=1}^{S} C_s$. The achievability part is proved in [116] using two different approaches:

1. **A random coding argument and a joint typicality decoding over product channels.** This coding scheme is based on the notion of product channels. Each possible permutation π yields a different product channel. Consequently, there are $S!$ possible product channels. These product channels have an input alphabet $\mathcal{X}^S$, and an output alphabet $\mathcal{Y}_1 \times \mathcal{Y}_2 \times \cdots \times \mathcal{Y}_S$. A random coding argument can be applied to each one of these product channels. Specifically, a properly chosen randomly code is shown to achieve the capacity C_{Π} under a joint-typicality decoding for all possible permutations π.

2. **A rate-matching coding scheme that is combined with a random coding argument, and a sequential joint-typicality decoding.** This coding scheme is based on the following concatenated structure: An information message $m \in [M]$ is mapped to a vector $\mathbf{m} = (m_1, m_2, \ldots, m_S)$ where $m_s \in [M^*]$ for

all $s \in [S]$ (this mapping is called a rate-matching code in [116]). It is assumed that $\frac{1}{n}\log_2 M^* < C^*$ where C^* is the maximal capacity of the S given parallel channels. Next, a randomly chosen codebook $\mathcal{C}$ with M^* codewords of block length n is chosen. Each element of $\mathbf{m}$ is encoded using the randomly chosen codebook $\mathcal{C}$, yielding a set of S codewords $\{\mathbf{x}_s\}_{s=1}^S$. The codewords $\{\mathbf{x}_s\}_{s=1}^S$ are transmitted over the considered parallel channels. The decoding procedure is based on a sequential joint-typicality decoding. First, the received vector over the channel with the maximal capacity is decoded using a standard joint typicality decoding. Next, the received vector over the channel with the second largest capacity is decoded using a joint-typical decoder. Over all possible codewords under this decoding rule, a message is chosen such that the rate-matching coding is satisfied. The decoding continues recursively, where in each decoding stage, a message is chosen such that the rate-matching code constraints are satisfied.

6.1.2 MDS codes

In this section some basic properties of MDS codes are provided. For complete details and proofs, the reader is referred to [74] and [89].

Definition 6.2 An (n, k) linear block code $\mathcal{C}$ whose minimum distance is d is called a maximum distance separable (MDS) code if

$$d = n - k + 1. \tag{6.4}$$

Remark 6.2 The RHS of (6.4) is the Singleton bound on the minimum distance of a linear block code.

Example 6.1 (MDS codes) The $(n, 1)$ repetition code, $(n, n-1)$ single parity-check (SPC) code, and the whole space of vectors over a finite field are all MDS codes.

The following properties of MDS codes are of interest in the continuation of this paper:

Proposition 6.1 (On the generator matrix of an MDS code) Let $\mathcal{C}$ be an MDS code of dimension k. Then, every k columns of the generator matrix of $\mathcal{C}$ are linearly independent.

Corollary 6.1 Every k symbols of a codeword in an MDS code of dimension k completely characterize the codeword.

Let $S > 0$ be an integer number and fix an integer $m > 0$ such that $2^m - 1 \geq S$. For all $k \in [2^m - 1]$, there exists a $(2^m - 1, k)$ RS code over the Galois field GF(2^m). Every RS code is an MDS code [89, Proposition 4.2]. In Section 6.3.3 a family of MDS codes with various block lengths and dimensions is applied to construct a parallel coding scheme. Two alternatives are suggested:

1. **Shortened RS codes:** Consider a $(2^m - 1, k)$ RS code over the Galois field GF(2^m). Deleting $2^m - 1 - S$ columns from the generator matrix of the considered code results in an (S, k) linear block code over the same alphabet. The resulting code is an (S, k) MDS code over GF(2^m).

2. **Generalized RS (GRS) codes:** GRS codes are MDS codes which can be constructed over GF(2^m) for every block length S and dimension k (as long as $2^m - 1 \geq S$).

Remark 6.3 (On the determination of codewords in RS and GRS codes) Our main interest in MDS codes is due to Corollay 6.1. This property is even more appealing for the case of RS or GRS codes because the determination of a codeword in RS or GRS codes is based on a polynomial interpolation over finite fields (see, e.g., [89, p. 151]).

6.2 Stochastically degraded parallel channels

The polarization properties of stochastically degraded parallel-channels are studied in this section.

Definition 6.3 (Stochastically degraded channels) Consider two memoryless channels with a common input alphabet $\mathcal{X}$, transition probability functions P_1 and P_2, and two output alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$, respectively. The channel P_2 is a stochastically degraded version of channel P_1 if there exists a channel D with an input alphabet $\mathcal{Y}_1$ and an output alphabet $\mathcal{Y}_2$ such that

$$P_2(y_2|x) = \sum_{y_1 \in \mathcal{Y}_1} P_1(y_1|x) D(y_2|y_1), \quad \forall x \in \mathcal{X}, y_2 \in \mathcal{Y}_2. \tag{6.5}$$

Lemma 6.1 (On the degradation of split channels) Let P_1 and P_2 be two transition probability functions with a common binary input alphabet $\mathcal{X} = \{0, 1\}$ and two output alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$, respectively. For a blocklength n, the split channels of P_1 and P_2 are denoted by $P_{1,n}^{(l)}$ and $P_{2,n}^{(l)}$, respectively, for all $l \in [n]$. Assume that the channel P_2 is a stochastically degraded version of channel P_1. Then, for every

$l \in [n]$, the split channel $P_{2,n}^{(l)}$ is a stochastically degraded version of the split channel $P_{1,n}^{(l)}$.

Proof: The proof follows by induction. First, the case of $n = 2$ is considered. For every $(y_{2,1}, y_{2,2}) \in \mathcal{Y}_2^2$ and $u_1 \in \mathcal{X}$:

$$\begin{aligned}
P_{2,2}^{(1)}(y_{2,1}, y_{2,2}|u_1) &\overset{(a)}{=} \sum_{u_2 \in \mathcal{X}} \frac{1}{2} P_2(y_{2,1}|u_1 + u_2) P_2(y_{2,2}|u_2) \\
&\overset{(b)}{=} \sum_{u_2 \in \mathcal{X}} \frac{1}{2} \left(\sum_{y_{1,1} \in \mathcal{Y}_1} P_1(y_{1,1}|u_1 + u_2) D(y_{2,1}|y_{1,1}) \right) \\
&\qquad \left(\sum_{y_{1,2} \in \mathcal{Y}_1} P_1(y_{1,2}|u_2) D(y_{2,2}|y_{1,2}) \right) \\
&= \frac{1}{2} \sum_{(y_{1,1}, y_{1,2}) \in \mathcal{Y}_1^2} D(y_{2,1}|y_{1,1}) D(y_{2,2}|y_{1,2}) \\
&\qquad \sum_{u_2 \in \mathcal{X}} P_1(y_{1,1}|u_1 + u_2) P_1(y_{1,2}|u_2) \\
&\overset{(c)}{=} \sum_{(y_{1,1}, y_{1,2}) \in \mathcal{Y}_1^2} D(y_{2,1}|y_{1,1}) D(y_{2,2}|y_{1,2}) P_{1,2}^{(1)}(y_{1,1}, y_{1,2}|u_1)
\end{aligned}$$

where (a) and (c) follow from (5.12), and (b) follows from (6.5). Hence, it is established that $P_{2,2}^{(1)}$ is a stochastically degraded version of $P_{1,2}^{(1)}$. Similar arguments verify that $P_{2,2}^{(2)}$ is a stochastically degraded version of $P_{1,2}^{(2)}$.

Next, assume that for $i > 1$, the split channel $P_{2,2^i}^{(l)}$ is a stochastically degraded version of the split channel $P_{1,2^i}^{(l)}$ for every $l \in [2^i]$. It is assumed that the degradation is with respect to the observations over the combined channel outputs. Specifically, it is assumed that

$$P_{2,2^i}^{(l)}(\mathbf{y}_2, \mathbf{u}|x) = \sum_{\mathbf{y}_1 \in \mathcal{Y}_1^n} D(\mathbf{y}_2|\mathbf{y}_1) P_{2,2^i}^{(l)}(\mathbf{y}_2, \mathbf{u}|x) \tag{6.6}$$

for every $l \in [2^i]$, $\mathbf{y}_2 \in \mathcal{Y}_2^{2^i}$, $\mathbf{u} \in \mathcal{X}^{l-1}$, and $x \in \mathcal{X}$. It follows that for every $l \in [2^i]$:

$$\begin{aligned}
P_{2,2^{i+1}}^{(2l-1)}\big((\mathbf{y}_2^{(1)}, \mathbf{y}_2^{(2)}), \mathbf{u}|x\big) &\overset{(a)}{=} \sum_{u \in \mathcal{X}} \frac{1}{2} P_{2,2^i}^{(l)}\big(\mathbf{y}_2^{(1)}, g(\mathbf{u})|x + u\big) P_{2,2^i}^{(l)}\big(\mathbf{y}_2^{(2)}, e(\mathbf{u})|u\big) \\
&\overset{(b)}{=} \sum_{u \in \mathcal{X}} \left\{ \frac{1}{2} \left(\sum_{\mathbf{y}_1^{(1)} \in \mathcal{Y}_1^{2^i}} D(\mathbf{y}_2^{(1)}|\mathbf{y}_1^{(1)}) P_{1,2^i}^{(l)}\big(\mathbf{y}_1^{(1)}, g(\mathbf{u})|x + u\big) \right) \right. \\
&\qquad \left. \left(\sum_{\mathbf{y}_1^{(2)} \in \mathcal{Y}_1^{2^i}} D(\mathbf{y}_2^{(2)}|\mathbf{y}_1^{(2)}) P_{2,2^i}^{(l)}\big(\mathbf{y}_1^{(2)}, e(\mathbf{u})|u\big) \right) \right\}
\end{aligned}$$

$$= \sum_{(\mathbf{y}_1^{(1)},\mathbf{y}_1^{(2)})\in\mathcal{Y}_1^{2^{i+1}}} \Bigg\{ D(\mathbf{y}_2^{(1)}|\mathbf{y}_1^{(1)}) D(\mathbf{y}_2^{(2)}|\mathbf{y}_1^{(2)}) \sum_{u\in\mathcal{X}} \frac{1}{2} P_{1,2^i}^{(l)}\big(\mathbf{y}_1^{(1)}, g(\mathbf{u})|x+u\big) P_{1,2^i}^{(l)}\big(\mathbf{y}_1^{(2)}, e(\mathbf{u})|u\big) \Bigg\}$$

$$\overset{(c)}{=} \sum_{(\mathbf{y}_1^{(1)},\mathbf{y}_1^{(2)})\in\mathcal{Y}_1^{2^{i+1}}} D(\mathbf{y}_2^{(1)}|\mathbf{y}_1^{(1)}) D(\mathbf{y}_2^{(2)}|\mathbf{y}_1^{(2)}) P_{1,2^{i+1}}^{(2l-1)}\big((\mathbf{y}_2^{(1)},\mathbf{y}_2^{(2)}),\mathbf{u}|x\big)$$

where $\mathbf{y}_2^{(1)}, \mathbf{y}_2^{(2)} \in \mathcal{Y}^{2^i}$, $\mathbf{u} \in \mathcal{X}^{2l-2}$, $x \in \mathbf{X}$, and the mappings g and e are defined in (5.14) and (5.15), respectively. The transitions in (a) and (c) follow from (5.12), and (b) follows from (6.6). Consequently, the split channel $P_{2,2^{i+1}}^{(2l-1)}$ is a stochastically degraded version of the split channel $P_{1,2^{i+1}}^{(2l-1)}$ for every $l \in [2^i]$. Similar arguments verify that for every $l \in [2^i]$ the split channel $P_{2,2^{i+1}}^{(2l)}$ is a stochastically degraded version of the split channel $P_{1,2^{i+1}}^{(2l)}$. Moreover, the degradation is with respect to the combined-channel observations in (6.6). ■

Remark 6.4 Note that the output alphabets of the split channels $P_{1,n}^{(l)}$ and $P_{2,n}^{(l)}$ are $\mathcal{Y}_1^n \times \mathcal{X}^{l-1}$ and $\mathcal{Y}_2^n \times \mathcal{X}^{l-1}$, respectively. In the proof of Lemma 6.1, a particular degradation is shown, which is with respect to the received vectors over the original channels (over the alphabets $\mathcal{Y}_1^n$ and $\mathcal{Y}_2^n$) where the split channel observations over $\mathcal{X}^{l-1}$ are left unaltered.

Definition 6.4 (Stochastically degraded parallel channels) Let $\{P_s\}_{s=1}^S$ be a set of S parallel memoryless channels, and denote the capacity of P_s by C_s for all $s \in [S]$. In addition, assume without loss of generality that $C_s \geq C_{s'}$ for all $1 \leq s < s' \leq S$. The channels $\{P_s\}_{s=1}^S$ are stochastically degraded if for every $1 \leq s < s' \leq S$ the channel $P_{s'}$ is a stochastically degraded version of P_s.

The following corollary is an application of Theorem 5.3 for a set of (stochastically) degraded parallel channels:

Corollary 6.2 (On monotonic information sets for stochastically degraded parallel channels) Consider a set of S memoryless degraded and symmetric parallel channels $\{P_s\}_{s=1}^S$, with a common binary-input alphabet $\mathcal{X}$. For every $s \in [S]$, denote the capacity of the channel P_s by C_s, and assume without loss of generality that

$$C_1 \geq C_2 \geq \cdots \geq C_S.$$

Fix $0 < \beta \leq \frac{1}{2}$ and a set of rates $\{R_s\}_{s=1}^S$ where

$$0 \leq R_s \leq C_s, \quad \forall s \in [S].$$

Then, there exists a sequence of information sets $\mathcal{A}_n^{(s)} \subseteq [n]$, $s \in [S]$ and $n = 2^i$ where $i \in \mathbb{N}$, satisfying the following properties:

1. Rate:
$$|\mathcal{A}_n^{(s)}| \geq nR_s, \quad \forall s \in [S]. \tag{6.7}$$

2. Monotonicity:
$$\mathcal{A}_n^{(S)} \subseteq \mathcal{A}_n^{(S-1)} \subseteq \cdots \subseteq \mathcal{A}_n^{(1)}. \tag{6.8}$$

3. Performance:
$$\Pr\big(\mathcal{E}_l(P_s)\big) \leq 2^{-n^\beta} \tag{6.9}$$
for all $l \in \mathcal{A}_n^{(s)}$ and $s \in [S]$, where $\mathcal{E}_l(P_s)$ is the error event defined in (5.16).

Proof: The rate and performance properties form immediate consequences of Theorem 5.3 and Proposition 5.3. Nevertheless, it is required to prove that the choice of the information set sequences can be made such that the monotonicity property in (6.8) is satisfied. Start with $s = S$. From Theorem 5.3 and Proposition 5.3, it follows that there exists a sequence of sets $\{\mathcal{A}_n^{(S)}\}$ satisfying (6.7) and (6.9). Next, fix an $s' \in [S]$ and assume that for all $s > s'$, the set sequences $\{\mathcal{A}_n^{(s)}\}$ can be chosen such that the properties in (6.7) and (6.9) are satisfied, and in addition

$$\mathcal{A}_n^{(S)} \subseteq \mathcal{A}_n^{(S-1)} \subseteq \cdots \subseteq \mathcal{A}_n^{(s'+1)}. \tag{6.10}$$

If $s' = S$ then (6.10) is satisfied in void. The existence of the sequence $\{\mathcal{A}_n^{(s')}\}$ satisfying (6.7) and (6.9) is already provided by Theorem 5.3 and Proposition 5.3. It is left to verify that the set sequence can be chosen such that the monotonicity property

$$\mathcal{A}_n^{(s'+1)} \subseteq \mathcal{A}_n^{(s')} \tag{6.11}$$

is kept. Choose an arbitrary index $l \in \mathcal{A}_n^{(s'+1)}$. It is proved that this index corresponds to the information set for the channel $P_{s'}$. Specifically, the performance property in (6.9) is satisfied for $s = s'$. Since $P_{s'+1}$ is a degraded version of $P_{s'}$, then according to Lemma 6.1, the split channel $P_{s'+1,n}^{(l)}$ is a degraded version of the split channel $P_{s',n}^{(l)}$. It is clearly suboptimal to first degrade the observation vector $\mathbf{y} \in \mathcal{Y}_{s'}$ to create a vector $\tilde{\mathbf{y}} \in \mathcal{Y}_{s'+1}$, and only then detect the input bit x for the degraded split channel. However, the detection error event for the degraded split channel $P_{s'+1,n}^{(l)}$ satisfies the upper bound in (6.9). As a result, the optimal detection error for the better split channel $P_{s',n}^{(l)}$ must also satisfy (6.9). Hence, all the indices in $\mathcal{A}_n^{(s'+1)}$ can be chosen for the set $\mathcal{A}_n^{(s')}$. The rest of indices are chosen arbitrarily out of the set of possible indices whose existence is guaranteed by Theorem 5.3. This establishes (6.11), and the proof follows by induction. ■

Remark 6.5 On good indices for stochastically degraded channels In Corollary 6.2, the existence of a monotonic sequence of information sets is proved for a degraded set of channels. A subtle inspection of the proof shows that the choice of the monotonic sequence of sets can be carried sequentially. First, the information set of the worst channel is specified. Then, as is shown in (6.11), all the indices that are "good" for the worse channel, are also "good" for the better channel. Here "good" is in the sense that the corresponding Bhattacharyya constants of the split channels (which form upper bounds on the corresponding decoding error probability) can be made exponentially low as the block length increases. Consequently, all that is left to specify are the rest of the "good" indices for the better channel (which are "not good" for the worse). The construction then follows sequentially.

Remark 6.6 Under the assumptions in Corollary 6.2, the capacity C_s for each of the channels in $\{P_s\}_{s=1}^{S}$ is achieved with equiprobable inputs. In cases where the parallel channels are not symmetric, a similar result can be shown where the capacities are replaced with the mutual information obtained with equiprobable inputs.

6.3 The Proposed Coding Scheme

In this section, parallel polar coding scheme is provided for a set of binary-input, memoryless, degraded, and symmetric parallel channels. First, two simple particular cases are studied in Sections 6.3.1 and 6.3.2 where transmission over $S = 2$ and 3 parallel channels is considered. Next, the general case is studied in Sections 6.3.3-6.3.4.

6.3.1 Parallel polar coding for $S = 2$ channels

The case of two memoryless degraded and symmetric parallel channels P_1 and P_2, whose capacities are equal (i.e., $C = C_1 = C_2$), is first considered. Fix a rate $R < C$, and choose a polar code for the channel P_2 at rate R. Denote the information index set for the chosen code by $\mathcal{A}_n^{(2)}$. According to Corollary 6.2, the same polar code is suitable for the channel P_1. Both polar codes are used with the same predetermined bit vector $\mathbf{b}$ for the indices in $[n] \setminus \mathcal{A}_n^{(2)}$. As both channels are symmetric, the particular choice of $\mathbf{b}$ can be made arbitrary for these channels. Assume that $2|\mathcal{A}_n^{(2)}|$ information bits are encoded by the two polar codes where $|\mathcal{A}_n^{(2)}|$ bits are encoded by each code. The particular assignment of information bits to polar codes can be chosen arbitrarily. Let $\mathbf{x}_1$ and $\mathbf{x}_2$ be the resulting codewords. Since the same polar code is used for both channels, the mapping of codewords to channels is not relevant

in this case and the standard successive decoding procedure in [4] can decode both $\mathbf{x}_1$ and $\mathbf{x}_2$, irrespectively of the channel assignments (as long as both codes use the same predetermined and fixed bits). The overall rate for the described scheme is $\frac{2}{n}|\mathcal{A}_n^{(2)}|$. As $R < C$ can be chosen arbitrarily close to C, all rates below $C_{\Pi} = 2C$ are achievable.

The case where $C_1 > C_2$ is addressed in the following. Set arbitrary rates $R_1 < C_1$ and $R_2 < C_2$, and construct a polar code for the degraded channel P_2. The polar code for P_2 is defined using the information index set $\mathcal{A}_n^{(2)}$ (where the information bits are assigned) and the predetermined and fixed bits assigned for the remaining indices in $[n] \setminus \mathcal{A}_n^{(2)}$. The size of the information set satisfies $|\mathcal{A}_n^{(2)}| \geq nR_2$. In contrast to polar codes used for a single channel, in the case of polar codes designed for transmission over $S = 2$ parallel channels, not all the symbols corresponding to indices from the set $[n] \setminus \mathcal{A}_n^{(2)}$ are assigned as predetermined and fixed bits. According to Corollary 6.2, the choice of $\mathcal{A}_n^{(1)}$ can be made such that $\mathcal{A}_n^{(2)} \subseteq \mathcal{A}_n^{(1)}$ and $|\mathcal{A}_n^{(1)}| \geq nR_1$. As the case of equal size information set can be treated similarly to the case above with $C_1 = C_2$, it is assumed that $\mathcal{A}_n^{(2)} \subset \mathcal{A}_n^{(1)}$ (i.e., a strict inclusion is assumed). For the indices specified by $[n] \setminus \mathcal{A}_n^{(1)}$, predetermined and fixed bits

$$\mathbf{b} \in \mathcal{X}^{n-|\mathcal{A}_n^{(1)}|} \tag{6.12}$$

are chosen in the recursive construction of the codeword $\mathbf{x}_1$. The vector $\mathbf{b}$ is also used for the same indices in the recursive construction of the second polar codeword $\mathbf{x}_2$, i.e., the indices in $[n] \setminus \mathcal{A}_n^{(1)}$ where

$$[n] \setminus \mathcal{A}_n^{(1)} \subset [n] \setminus \mathcal{A}_n^{(2)}.$$

It is left to determine the status of the bits corresponding to the the indices in the set

$$\left([n] \setminus \mathcal{A}_n^{(2)}\right) \setminus \left([n] \setminus \mathcal{A}_n^{(1)}\right) = \mathcal{A}_n^{(1)} \setminus \mathcal{A}_n^{(2)} \tag{6.13}$$

for the construction of the second codeword. Note that if the design of polar codes for a single channel is considered, the indices in (6.13) are information indices for the polar code designed for the channel P_1, but they should correspond to predetermined and fixed bit indices for polar coding over the channel P_2. For parallel polar coding, on the other hand, the bits corresponding to the indices in (6.13) for the second codeword are set to be the same as the information bits encoded by the first codeword. Therefore, the set of bits for the recursive construction in the indices of (6.13) are called the repetition bits.

To describe the encoding procedure in terms of coset coding, recall the equivalence between the recursive construction and coset coding as stated in (5.7). Let $k_1 = |\mathcal{A}_n^{(1)}|$

and $k_2 = |\mathcal{A}_n^{(2)}|$. Denote the information message bits by $\mathbf{u}_1 \in \mathcal{X}^{k_2}$, $\mathbf{u}_2 \in \mathcal{X}^{k_2}$ and $\mathbf{u}_{\rm r} \in \mathcal{X}^{k_1-k_2}$. The vector $\mathbf{u}_{\rm r}$ includes the repetition bits of the parallel polar construction. The first codeword is given by

$$\mathbf{x}_1 = \mathbf{u}_1 G_n\left(\mathcal{A}_n^{(2)}\right) + \mathbf{u}_{\rm r} G_n\left(\mathcal{A}_n^{(1)} \setminus \mathcal{A}_n^{(2)}\right) + \mathbf{b} G_n\left([n] \setminus \mathcal{A}_n^{(1)}\right) \tag{6.14}$$

where $\mathbf{b}$ designates the vector of the predetermined and fixed bits in (6.12), and G_n is the polar generator matrix. Note that $\mathbf{x}_1$ satisfies

$$\mathbf{x}_1 = m_1(\mathbf{u}_1, \mathbf{u}_{\rm r}) \cdot G_n\left(\mathcal{A}_n^{(1)}\right) + \mathbf{b} G_n\left([n] \setminus \mathcal{A}_n^{(1)}\right)$$

where $m_1(\mathbf{u}_1, \mathbf{u}_{\rm r}) \in \mathcal{X}^{k_1}$ is a proper permutation of the vector $(\mathbf{u}_1, \mathbf{u}_{\rm r})$. Therefore, the definition in (6.14) is equivalent to the polar coding in (5.7). The second codeword is given by

$$\mathbf{x}_2 = \mathbf{u}_2 G_n\left(\mathcal{A}_n^{(2)}\right) + \mathbf{u}_{\rm r} G_n\left(\mathcal{A}_n^{(1)} \setminus \mathcal{A}_n^{(2)}\right) + \mathbf{b} G_n\left([n] \setminus \mathcal{A}_n^{(1)}\right). \tag{6.15}$$

As mentioned, this is almost like a standard polar encoding where the difference is that some of the predetermined and fixed bits form a repetition of information bits. In fact, the second codeword can be written by

$$\mathbf{x}_2 = \mathbf{u}_2 G_n\left(\mathcal{A}_n^{(2)}\right) + m_2(\mathbf{u}_{\rm r}, \mathbf{b}) \cdot G_n\left([n] \setminus \mathcal{A}_n^{(2)}\right)$$

where $m_2(\mathbf{u}_{\rm r}, \mathbf{b})$ is a proper permutation of the vector $(\mathbf{u}_{\rm r}, \mathbf{b})$.

The decoding starts with the channel P_1 whose capacity is maximal ($C_1 > C_2$). No matter what the actual codeword is transmitted over P_1 (either $\mathbf{x}_1$ or $\mathbf{x}_2$), a standard polar successive cancellation decoding procedure is applied to decode the set of information bits corresponding to the indices in $\mathcal{A}_n^{(1)}$. If $\mathbf{x}_1$ is the codeword transmitted over P_1, then the bit vectors $\mathbf{u}_1$ and $\mathbf{u}_{\rm r}$ are decoded. Else, if the codeword $\mathbf{x}_2$ is transmitted over the channel P_1, then $\mathbf{u}_2$ and $\mathbf{u}_{\rm r}$ are decoded. Next, recall that the vector $\mathbf{b}$ is predetermined and fixed. In addition, the repetition bit vector $\mathbf{u}_{\rm r}$, corresponding to the indices in (6.13), are already decoded after the previous decoding step. Note that the vector $\mathbf{u}_{\rm r}$ is available after the first decoding step irrespectively of the actual transmission assignment of the codewords $\mathbf{x}_1$ and $\mathbf{x}_2$ to the parallel channels. Hence, using the repetition bits $\mathbf{u}_{\rm r}$ as if they are predetermined, the successive cancellation decoding can be applied to the channel P_2. If $\mathbf{x}_1$ is transmitted over P_1 and $\mathbf{x}_2$ is transmitted over P_2, then the bits $\mathbf{u}_2$ are decoded using the successive cancellation decoding where $m_2(\mathbf{u}_{\rm r}, \mathbf{b})$ are used as predetermined and fixed bits. Otherwise, if $\mathbf{x}_2$ is transmitted over P_1 and $\mathbf{x}_1$ is transmitted over P_2, then the bits $\mathbf{u}_1$ are decoded using the successive cancellation decoding where again $m_2(\mathbf{u}_{\rm r}, \mathbf{b})$ are used as predetermined and fixed bits. This completes the decoding of all the information bits. As $R_1 < C_1$ and $R_2 < C_2$ can be chosen arbitrarily close to C_1 and C_2, respectively, then the transmission rate $C_1 + C_2$ is achievable.

6.3.2 Parallel polar coding for $S = 3$ channels

Assume that a parallel coding scheme is applied for communication over a set of three parallel channels P_1, P_2, and P_3, whose capacities are $C_1 > C_2 > C_3$, respectively. According to Theorem 6.1, the capacity C_{Π} in this case satisfies

$$C_{\Pi} = C_1 + C_2 + C_3.$$

Fix the rates $R_1 > R_2 > R_3$, satisfying $R_s < C_s$ for all $s \in [3]$, and let

$$R \triangleq R_1 + R_2 + R_3.$$

In the following, a parallel polar coding scheme of rate R is described that achieves reliable communication. Therefore, the proposed scheme achieves the capacity C_{Π} by selecting the rates R_1, R_2, and R_3 to be close, respectively, to C_1, C_2, and C_3, and satisfy the above condition for the rate triple.

Let $\{\mathcal{A}_n^{(s)}\}$ be the information set sequences as in Corollary 6.2. Fix a block length n, let

$$k_s \triangleq |\mathcal{A}_n^{(s)}|, \quad s \in [3]$$

and

$$k \triangleq k_1 + k_2 + k_3.$$

The encoding of k information bits to 3 codewords: $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ is defined. First, the information bits are arbitrarily partitioned into three groups of sizes k_1, k_2 and k_3. Next, the encoding of the first two codewords is performed as follows:

- The k_1 information bits used to encode $\mathbf{x}_1$ are (arbitrarily) partitioned to three subsets: $\mathbf{u}_{1,1} \in \mathcal{X}^{k_3}$, $\mathbf{u}_{1,2} \in \mathcal{X}^{k_2-k_3}$, and $\mathbf{u}_{\mathrm{r}} \in \mathcal{X}^{k_1-k_2}$.

- The k_2 information bits used to encode $\mathbf{x}_2$ are (arbitrarily) partitioned into two subsets: $\mathbf{u}_{2,1} \in \mathcal{X}^{k_3}$ and $\mathbf{u}_{2,2} \in \mathcal{X}^{k_2-k_3}$. In addition, $\mathbf{u}_{\mathrm{r}}$ (used for encoding $\mathbf{x}_1$) is also involved in the encoding of $\mathbf{x}_2$.

- The codewords $\mathbf{x}_1$ and $\mathbf{x}_2$ are defined similarly to the case of $S = 2$ parallel channels. Specifically, in terms of coset codes:

$$\begin{aligned}\mathbf{x}_1 = {} & \mathbf{u}_{1,1}G_n\left(\mathcal{A}_n^{(3)}\right) + \mathbf{u}_{1,2}G_n\left(\mathcal{A}_n^{(2)} \setminus \mathcal{A}_n^{(3)}\right) \\ & + \mathbf{u}_{\mathrm{r}}G_n\left(\mathcal{A}_n^{(1)} \setminus \mathcal{A}_n^{(2)}\right) + \mathbf{b}G_b\left([n] \setminus \mathcal{A}_n^{(1)}\right) \qquad (6.16) \\ \mathbf{x}_2 = {} & \mathbf{u}_{2,1}G_n\left(\mathcal{A}_n^{(3)}\right) + \mathbf{u}_{2,2}G_n\left(\mathcal{A}_n^{(2)} \setminus \mathcal{A}_n^{(3)}\right) \\ & + \mathbf{u}_{\mathrm{r}}G_n\left(\mathcal{A}_n^{(1)} \setminus \mathcal{A}_n^{(2)}\right) + \mathbf{b}G_b\left([n] \setminus \mathcal{A}_n^{(1)}\right) \qquad (6.17)\end{aligned}$$

  where $\mathbf{b} \in \mathcal{X}^{n-k}$ is a predetermined and fixed vector.

The encoding of the codeword $\mathbf{x}_3$ is based on the remaining k_3 information bits, denoted by $\mathbf{u}_3 \in \mathcal{X}^{k_3}$. In addition, the information bits in $\mathbf{u}_{1,2}$, $\mathbf{u}_{2,2}$ and $\mathbf{u}_{\mathrm{r}}$ are also involved in the encoding of $\mathbf{x}_3$:

$$\begin{aligned}\mathbf{x}_3 = \mathbf{u}_3 G_n\left(\mathcal{A}_n^{(3)}\right) + \left(\mathbf{u}_{1,2} + \mathbf{u}_{2,2}\right) G_n\left(\mathcal{A}_n^{(2)} \setminus \mathcal{A}_n^{(3)}\right) \\ + \mathbf{u}_{\mathrm{r}} G_n\left(\mathcal{A}_n^{(1)} \setminus \mathcal{A}_n^{(2)}\right) + \mathbf{b} G_n\left([n] \setminus \mathcal{A}_n^{(1)}\right).\end{aligned}$$

Note that the repetition approach is also done for the indices in $[n] \setminus \mathcal{A}_n^{(2)}$. However, a different approach is applied to the indices in $\mathcal{A}_n^{(2)} \setminus \mathcal{A}_n^{(3)}$. The bits corresponding to these indices are set using a symbol-wise parity-check of $\mathbf{u}_{1,2}$ and $\mathbf{u}_{2,2}$.

The order of decoding the information bits for all possible assignments of codewords over a set of three parallel channels is provided in Table 6.1. The decoding starts with the channel P_1 with the maximal capacity C_1. Irrespectively of the actual codeword that is transmitted over P_1, the bits which correspond to the indices in $\mathcal{A}_n^{(1)}$ are decoded using the standard polar successive cancellation decoding. The decoded bits depend on the actual codeword which is transmitted over P_1. Next, the decoding proceeds to process the vector observed at the output of the channel P_2, whose capacity is C_2. The decoding of $|\mathcal{A}_n^{(2)}|$ information bits is established in this decoding step. Note that for a standard successive cancellation decoding procedure, $n - |\mathcal{A}_n^{(2)}|$ predetermined and fixed bits are required for proper operation. For the case at hand, these bits are not all predetermined and fixed. The vector $\mathbf{b}$ is predetermined, but the rest depends on the repetition bits $\mathbf{u}_{\mathrm{r}}$. Since the bits $\mathbf{u}_{\mathrm{r}}$ were decoded at the previous decoding stage (based on the observation vector of P_1), they can be treated as if they are predetermined and fixed for the decoding of $\mathbf{x}_2$. Consequently, $|\mathcal{A}_n^{(2)}|$ information bits are decoded (depending on the actual codeword transmitted over the channel P_2). Finally, the decoding proceeds for the vector received at the output of the channel P_3. As in the previous decoding steps, the polar successive cancellation decoding is applied where the bits corresponding to the split channels indexed by $[n] \setminus \mathcal{A}_n^{(3)}$ are not all predetermined and fixed (as in contrast to the standard single channel case). Nevertheless, these bits can be all determined using the information bits decoded in the two first steps. The bits in $\mathbf{b}$ are predetermined and fixed. The repetition bits in $\mathbf{u}_{\mathrm{r}}$ are already available after the decoding of the information transmitted over P_1. The rest, can be evaluated by taking a bit-wise exclusive-or (xor) of the bits decoded in the two previous steps. As an example, a combination shown in Table 6.1 is described explicitly. Consider the case where the codeword $\mathbf{x}_2$ is transmitted over the channel P_1, and the codeword $\mathbf{x}_3$ is transmitted over the channel P_2. At the first decoding step, the vectors $\mathbf{u}_{2,1}$, $\mathbf{u}_{2,2}$ and $\mathbf{u}_{\mathrm{r}}$ are decoded (where the predetermined bits refer to the vector $\mathbf{b}$). Next, the vectors $\mathbf{u}_3$, and $\mathbf{u}_{1,2} + \mathbf{u}_{2,2}$, are decoded (the

Channel P_1		Channel P_2		Channel P_3	
Transmitted Codeword	**Decoded Information**	**Transmitted Codeword**	**Decoded Information**	**Transmitted Codeword**	**Decoded Information**
$\mathbf{x}_1$	$\mathbf{u}_{1,1}$, $\mathbf{u}_{1,2}$, $\mathbf{u}_\mathrm{r}$	$\mathbf{x}_2$	$\mathbf{u}_{2,1}$, $\mathbf{u}_{2,2}$	$\mathbf{x}_3$	$\mathbf{u}_3$
		$\mathbf{x}_3$	$\mathbf{u}_3$, $\mathbf{u}_{1,2}+\mathbf{u}_{2,2}$	$\mathbf{x}_2$	$\mathbf{u}_{2,1}$
$\mathbf{x}_2$	$\mathbf{u}_{2,1}$, $\mathbf{u}_{2,2}$, $\mathbf{u}_\mathrm{r}$	$\mathbf{x}_1$	$\mathbf{u}_{1,1}$, $\mathbf{u}_{1,2}$	$\mathbf{x}_3$	$\mathbf{u}_3$
		$\mathbf{x}_3$	$\mathbf{u}_3$, $\mathbf{u}_{1,2}+\mathbf{u}_{2,2}$	$\mathbf{x}_1$	$\mathbf{u}_{1,1}$
$\mathbf{x}_3$	$\mathbf{u}_3$, $\mathbf{u}_{1,2}+\mathbf{u}_{2,2}$, $\mathbf{u}_\mathrm{r}$	$\mathbf{x}_1$	$\mathbf{u}_{1,1}$, $\mathbf{u}_{1,2}$	$\mathbf{x}_2$	$\mathbf{u}_{2,1}$
		$\mathbf{x}_2$	$\mathbf{u}_{2,1}$, $\mathbf{u}_{2,2}$	$\mathbf{x}_1$	$\mathbf{u}_{1,1}$

Table 6.1: The order of decoding the information bits for all possible assignment of codewords over a set of three parallel channels

pretermitted bits for this decoding stage refer to $\mathbf{b}$ and $\mathbf{u}_\mathrm{r}$). After this stage, the information bits $\mathbf{u}_{1,2}$ can be determined by

$$\mathbf{u}_{2,2} + (\mathbf{u}_{1,2} + \mathbf{u}_{2,2}) .$$

Moreover, the information bits $\mathbf{u}_{1,2}$ are used for the last decoding stage as predetermined and fixed bits (together with the vectors $\mathbf{u}_\mathrm{r}$ and $\mathbf{b}$). After the last decoding stage the vector $\mathbf{u}_{1,1}$ is decoded, and the decoding of all the information bits is completed.

To complete the current discussion, the case where some of the channels have equal capacities is concerned. One option is the trivial case where $C_1 = C_2 = C_3$. For this case, a regular polar encoding and decoding is applied. As long as the channels are degraded and symmetric, the information index sets and the predetermined and fixed bits are the same for all transmitted codewords. Consequently, irrespective of the selected permutation at the transmission, all the information bits can be decoded. The treatment of the case $C_1 > C_2 = C_3$ can be treated in a similar fashion to the case of $S = 2$ parallel channels. For the case where $C_1 = C_2 > C_3$, the parity-check construction should be applied.

6.3.3 Parallel polar coding for $S > 3$ channels

C.1. Encoding

A parallel polar encoding is described for the general case. The technique used for rate-matching encoding in [116] is incorporated in the current case as well. This technique is based on MDS codes, in particular (punctured) RS codes are used in [116] for rate splitting. As commented in Section 6.1.2, GRS codes can also fit for the provided construction. A set of $S-1$ MDS codes over the Galois field GF(2^m), all with a common block length S are chosen (either by puncturing an appropriate RS code or using GRS codes). These codes are denoted by $\mathcal{C}_{\mathrm{MDS}}^{(k)}$, $k \in [S-1]$, where the code $\mathcal{C}_{\mathrm{MDS}}^{(k)}$ has dimension k.

Let $\{P_s\}_{s=1}^S$ be a given set of memoryless degraded and symmetric parallel channels, whose capacities are ordered such that $C_1 > C_2 > \cdots > C_S$. Let $\{\mathcal{A}_n^{(s)}\}_{s=1}^S$ be the information index sets satisfying the properties in Corollary 6.2, for a block length n and rates $R_1 > R_2 > \cdots > R_S$, $R_s < C_s$, $s \in [S]$. Define

$$k_s \triangleq |\mathcal{A}_n^{(s)}|, \quad s \in [S]$$

and

$$k_{S+1} \triangleq 0.$$

In addition, it is assumed that n and k_s for all $s \in [S]$, are integral multiples of m. In the provided coding scheme, $k = \sum_{s=1}^S k_s$ information bits are encoded into S codewords $\mathbf{x}_s$, $s \in [S]$. As the rates R_s, $s \in [S]$ can be chosen arbitrarily close to C_s, respectively, the capacity C_Π in (6.3) is shown to be asymptotically achievable (the error performance is considered in Section 6.3.4).

Prior to the stage of polar encoding, the k information bits are first mapped into a set of binary vectors

$$\mathcal{U} = \left\{\mathbf{u}_{s,l} \in \mathcal{X}^{k_{S-l+1}-k_{S-l+2}} : \quad s, l \in [S]\right\}.$$

The $S \cdot k_S$ bits in the vectors $\mathbf{u}_{s,1}$, $s \in [S]$ are plain information bits, chosen arbitrarily from the set of k information bits. The vector set

$$\mathcal{C}_2 \triangleq \left\{\mathbf{u}_{s,2} = \big(u_{s,2}(1), u_{s,2}(2), \ldots, u_{s,2}(k_{S-1} - k_S)\big) : \quad s \in [S-1]\right\}$$

are also filled with plain information bits, chosen arbitrarily from the set of remaining $k - S \cdot k_S$ information bits (note that under the above assumptions $k - S \cdot k_S > 0$). Next, the vector $\mathbf{u}_{S,2}$ is determined (the following steps are accompanied with the illustration in Figure 6.2):

1. Each vector in $\mathcal{C}_2$ is rewritten as a row vector of a matrix over GF(2^m) (this step is illustrated in Figure 6.2 where each vector is represented with a horizontal rectangle). Each m consecutive bits are mapped into a symbol over GF(2^m). This results in the $(S-1) \times K_{S-1,S}$ matrix over GF(2^m)

$$C^{(2)} = \left(C_{i,j}^{(2)}\right), \quad i \in [S-1],\ j \in [K_{S-1,S}]$$

where

$$K_{S-1,S} \triangleq \frac{k_{S-1} - k_S}{m}.$$

The element $C_{i,j}^{(2)}$ is the symbol over GF(2^m) corresponding to the binary length-m vector

$$\Big(\mathbf{u}_{i,2}\big((j-1)m+1\big), \mathbf{u}_{i,2}\big((j-1)m+2\big), \ldots, \mathbf{u}_{i,2}\big(jm\big)\Big)$$

where $i \in [S-1]$ and $j \in [K_{S-1,S}]$.

2. Each one of the columns of $C^{(2)}$ are considered as the first $S-1$ symbols of a codeword in the code $\mathcal{C}_{\text{MDS}}^{(S-1)}$. These columns are illustrated with dashed vertical rectangles in Figure 6.2. Consequently, these columns completely determine the codewords $\{\mathbf{c}_j : \quad j \in [K_{S-1,S}]\}$ in the MDS $[S, S-1]$ code $\mathcal{C}_{\text{MDS}}^{(S-1)}$.

3. A length-$K_{S-1,S}$ vector $\tilde{\mathbf{u}}_{S,2}$ over GF(2^m) is defined using the last symbol of each of the codewords $\mathbf{c}_j$, $j \in [K_{S-1,S}]$, evaluated in the last step. Each of these symbols is illustrated as a filled black square in Figure 6.2.

4. The vector $\mathbf{u}_{S,2}$ is defined by the binary representation of the vector $\tilde{\mathbf{u}}_{S,2}$ where each symbol over GF(2^m) is replaced by its corresponding binary length-m vector.

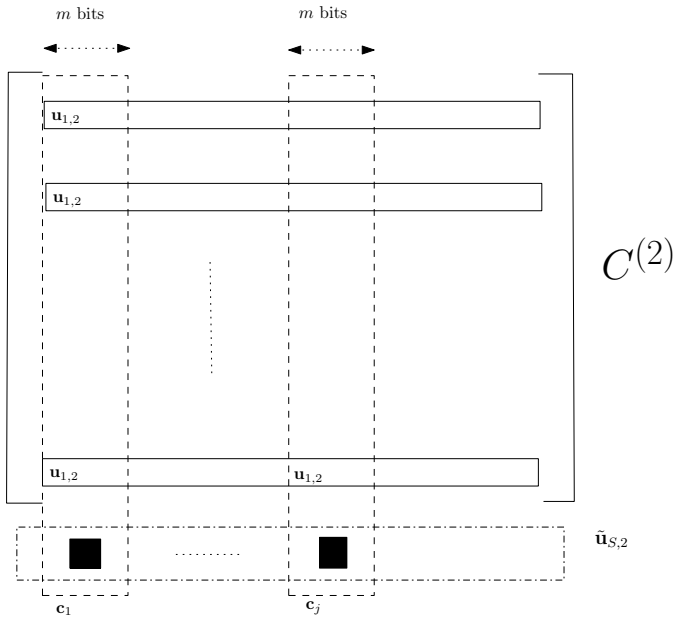


Figure 6.2: Illustration of the construction of the vector $\tilde{\mathbf{u}}_{S,2}$. The vectors $\mathbf{u}_{k,s}$, $k \in [S-1]$ defining the matrix $C^{(2)}$ are shown, along the columns defining the codewords $\mathbf{c}_j$, $j \in [K_{S-1,S}]$ in $\mathcal{C}_{\text{MDS}}^{(S-1)}$

The definition of the remaining vectors in $\mathcal{U}$ continues in a similar way. Let $2 < l \leq S$, and assume that the vectors $\mathbf{u}_{s,l'}$ are already defined for all $s \in [S]$ and $l' < l$, based on

$$\sum_{s=1}^{l'} (S-(s-1))(k_{S-(s-1)} - k_{S-(s-2)})$$

information bits (from a total of k information bits). The construction phase for the vectors $\mathbf{u}_{s,l}$, $s \in [S]$ is defined as follows:

1. The binary vector set

$$\mathcal{C}_l = \{\mathbf{u}_{s,l} : \ 1 \leq s \leq S-(l-1)\}$$

are filled with

$$\big(S-(l-1)\big)\left(k_{S-(l-1)} - k_{S-(l-2)}\right)$$

arbitrarily chosen information bits, out of the remaining

$$k - \sum_{s=1}^{l'} \big(S-(s-1)\big)\left(k_{S-(s-1)} - k_{S-(s-2)}\right)$$

information bits.

2. Each vector in $\mathcal{C}_l$ is rewritten over GF(2^m) as a row vector in an $(S-(l-1)) \times K_{S-(l-1),S-(l-2)}$ matrix over GF(2^m)

$$C^{(l)} = \left(C_{i,j}^{(l)}\right)$$

where

$$K_{S-(l-1),S-(l-2)} \triangleq \frac{k_{S-(l-1)} - k_{S-(l-2)}}{m}$$

and $C_{i,j}^{(l)}$, $i \in [S-(l-1)]$, $j \in [K_{S-(l-1),S-(l-2)}]$, equals the symbol in GF(2^m) corresponding to the binary length-m vector

$$\Big(\mathbf{u}_{i,l}\big((j-1)m+1\big), \mathbf{u}_{i,l}\big((j-1)m+2\big), \ldots, \mathbf{u}_{i,2}\big(jm\big)\Big).$$

3. Each column in C_l is a vector of $S-(l-1)$ symbols over GF(2^m). Hence, it completely determines a codeword $\mathbf{c}_j = (c_{j,1}, c_{j,2}, \ldots, c_{j,S})$, $j \in [K_{S-(l-1),S-(l-2)}]$, in the MDS $[S, S-(l-1)]$ code $\mathcal{C}_{\text{MDS}}^{(S-(l-1))}$. The columns of C_l are considered as the first $S-(l-1)$ symbols of a codeword in the code $\mathcal{C}_{\text{MDS}}^{(S-(l-1))}$.

4. Evaluate the remaining symbols for each of the codewords

$$\mathbf{c}_j, \quad j \in [K_{S-(l-1),S-(l-2)}].$$

5. The length-$K_{S-(l-1),S-(l-2)}$ vectors $\tilde{\mathbf{u}}_{s,l} = (\tilde{u}_{s,l}(1), \ldots, \tilde{u}_{s,l}(K_{S-(l-1),S-(l-2)}))$, $s > S-(l-1)$, over GF(2^m) are defined using the codewords $\mathbf{c}_j$, $j \in [K_{S-(l-1),S-(l-2)}]$ according to

$$\tilde{u}_{s,l}(j) = c_{j,s}.$$

6. For every $s > S-(l-1)$, The vector $\mathbf{u}_{s,l}$ is defined to be the binary representation of the vector $\tilde{\mathbf{u}}_{s,l}$ (where each symbol over GF(2^m) is replaced with its binary length-m vector representation).

The parallel polar codewords are defined using the coset code notation. Specifically, the codewords $\mathbf{x}_s$, $s \in [S]$, are defined according to

$$\mathbf{x}_s = \sum_{l=1}^{S} \mathbf{u}_{s,l} G_n \left(\mathcal{A}_n^{(S-(l-1))} \setminus \mathcal{A}_n^{(S-(l-2))} \right) + \mathbf{b} G_n \left([n] \setminus \mathcal{A}_n^{(1)} \right), \quad s \in [S] \tag{6.18}$$

where $\mathcal{A}_n^{(S+1)} \triangleq \emptyset$ and $\mathbf{b} \in \mathcal{X}^{n-k_1}$ is a binary predetermined and fixed vector.

C.2. Decoding

The decoding process starts with the observations received at the output of the channel P_1 whose capacity is maximal. Assume that the codeword $x_{\pi^{-1}(1)}$ is transmitted over P_1. A polar successive cancellation decoding, with respect to the information index set $\mathcal{A}_n^{(1)}$, is applied to the received vector. This allows the decoding of the vectors $\mathbf{u}_{\pi^{-1}(1),l}$, $l \in [S]$ (as if they are the information bits of the considered polar code). If $\pi^{-1}(1) = 1$, then indeed all the vectors $\mathbf{u}_{\pi^{-1}(1),l} = \mathbf{u}_{1,l}$, $l \in [S]$ are information bit vectors. Generally, only a subset of these vectors comprise of information bits, the rest are coded binary representation of coded symbols of the chosen MDS codes.

At the second stage, the decoding of the received vector over P_2, which denotes probability transition of the channel with the second largest capacity, is concerned. Assume that the codeword $\mathbf{x}_{\pi^{-1}(2)}$ is transmitted over P_2. A polar successive cancellation decoding is used. This decoding procedure is capable of decoding $|\mathcal{A}_n^{(2)}|$ bits based on $n - |\mathcal{A}_n^{(2)}|$ predetermined and fixed bits. For the current decoding procedure, $n - |\mathcal{A}_n^{(1)}|$ of these bits are the predetermined and fixed bits in $\mathbf{b}$. The rest of $|\mathcal{A}_n^{(1)}| - |\mathcal{A}_n^{(2)}|$ bits are based on the bits decoded at the previous decoding stage. Specifically, the bit vector $\mathbf{u}_{\pi^{-1}(2),S}$ can be evaluated using the bit vector $\mathbf{u}_{\pi^{-1}(1),S}$. Recall that $\mathbf{u}_{\pi^{-1}(2),S}$ is the binary representation of $\tilde{\mathbf{u}}_{\pi^{-1}(2),S}$. Moreover, each of the symbols of $\tilde{\mathbf{u}}_{\pi^{-1}(2),S}$ belongs to a codeword in the $[S, 1]$ MDS code $\mathcal{C}_{\text{MDS}}^{(1)}$. These codewords are fully determined from the vector $\mathbf{u}_{\pi^{-1}(1),S}$ as follows:

1. Rewrite the vector $\mathbf{u}_{\pi^{-1}(1),S}$ over $\text{GF}(2^m)$ where each consecutive m bits are rewritten by the corresponding symbol over $\text{GF}(2^m)$. Denote by

   $$\tilde{\mathbf{u}}_{\pi^{-1}(1),S} = \left(\tilde{u}_{\pi^{-1}(1),S}(1), \ldots, \tilde{u}_{\pi^{-1}(1),S}(K_{1,2}) \right)$$

   the resulting length-$K_{1,2}$ vector over $\text{GF}(2^m)$.

2. For each symbol $\tilde{u}_{\pi^{-1}(1),S}(j)$, $j \in [K_{1,2}]$, find the codeword

   $$\mathbf{c}_j = (c_{j,1}, \ldots, c_{j,S}) \in \mathcal{C}_{\text{MDS}}^{(1)}$$

   whose $\pi^{-1}(1)$-th symbol satisfies $c_{j,\pi^{-1}(1)} = \tilde{u}_{\pi^{-1}(1),S}(j)$. These codewords are fully determined by the considered symbols.

3. Define the vector

$$\tilde{\mathbf{u}}_{\pi^{-1}(2),S} = \big(\tilde{u}_{\pi^{-1}(2),S}(1), \ldots, \tilde{u}_{\pi^{-1}(2),S}(K_{1,2})\big)$$

according to $\tilde{u}_{\pi^{-1}(2),S}(j) = c_{j,\pi^{-1}(2)}$ for every $j \in [K_{1,2}]$.

4. The vector

$$\mathbf{u}_{\pi^{-1}(2),S} = \big(u_{\pi^{-1}(2),S}(1), \ldots, u_{\pi^{-1}(2),S}(k_1 - k_2)\big)$$

is set to the binary representation of $\tilde{\mathbf{u}}_{\pi^{-1}(2),S}$. That is, the bits

$$u_{\pi^{-1}(2),S}((j-1)m+1), \ldots, u_{\pi^{-1}(2),S}(jm)$$

are the binary representation of the symbol

$$\tilde{u}_{\pi^{-1}(2),S}(j) \in \mathrm{GF}(2^m), \quad j \in K_{1,2}.$$

With both $\mathbf{b}$ and $\mathbf{u}_{\pi^{-1}(2),S}$ as predetermined and fixed bits, the polar successive cancellation decoding can be applied. Consequently, after the second decoding stage, all the S binary vectors $\mathbf{u}_{\pi^{-1}(2),s}$, $s \in [S]$, are fully determined. Moreover, based on the codewords $\mathbf{c}_j$, $j \in [K_{1,2}]$, the vectors $\mathbf{u}_{\pi^{-1}(s),S}$, are fully determined for all $s \geq 2$ as well.

Next, the remaining $S-2$ decoding stages are described. It is assumed that after the $(s-1)$-st decoding stage, where $2 < s < S$, the vectors $\mathbf{u}_{\pi^{-1}(s'),l}$ for either $1 \leq s' < s$ and $l \in [S]$, or $s' \geq s$ and $S - s + 3 \leq l \leq S$, were decoded at previous stages. At the s-th stage, the decoding is extended for the vectors $\mathbf{u}_{\pi^{-1}(s),l}$ for all $l \in [S]$ and the vectors $\mathbf{u}_{\pi^{-1}(s'),S-s+2}$ for all $s' \in [S]$.

In order to apply the polar successive cancellation decoding procedure to the vector received over the channel P_s, the bits in $\mathbf{b}$ and $\{\mathbf{u}_{\pi^{-1}(s),l}\}_{l \geq S-(s-2)}$ must be known for the procedure. The vector $\mathbf{b}$ is clearly known. In addition, the bits in $\{\mathbf{u}_{\pi^{-1}(s),l}\}_{l \geq S-(s-3)}$ are already decoded in previous stages. It is left to determine the bits in $\mathbf{u}_{\pi^{-1}(s),S-(s-2)}$. These bits are determined in a similar manner as in the decoding stage for $s = 2$, where the vector $\mathbf{u}_{\pi^{-1}(2),S}$ is determined. Moreover, the determination of $\mathbf{u}_{\pi^{-1}(s),S-(s-2)}$ is established along with the determination of $\mathbf{u}_{\pi^{-1}(s'),S-(s-2)}$ for all $s' \geq s$, in the following way:

1. The binary vectors $\mathbf{u}_{\pi^{-1}(s'),S-s+2}$ for $s' < s$ are already decoded at previous stages. Rewrite these vectors over $\mathrm{GF}(2^m)$ where each consecutive m bits are rewritten by the corresponding symbol over $\mathrm{GF}(2^m)$. Denote the set of resulting vectors by

$$\mathcal{D} = \left\{\tilde{\mathbf{u}}_{\pi^{-1}(s'),S-s+2} = \big(\tilde{u}_{\pi^{-1}(s'),S-s+2}(1), \ldots, \tilde{u}_{\pi^{-1}(s'),S-s+2}(K_{s-1,s})\big) : \quad s' < s\right\}.$$

2. The set $\mathcal{D}$ completely describes $K_{s-1,s}$ codeword

$$\mathbf{c}_j = (c_{j,1}, \ldots, c_{j,S}), \quad j \in [K_{s-1,s}]$$

all in the code $\mathcal{C}_{\text{MDS}}^{(s-1)}$ and satisfy the constraints:

$$c_{j,\pi^{-1}(s')} = \tilde{u}_{\pi^{-1}(s'),S-s+2}(j), \quad 1 \leq s' < s. \tag{6.19}$$

3. Define the vectors

$$\tilde{\mathbf{u}}_{\pi^{-1}(s'),S-s+2} = \left(\tilde{u}_{\pi^{-1}(s'),S-s+2}(1), \ldots, \tilde{u}_{\pi^{-1}(s'),S-s+2}(K_{s-1,s})\right)$$

for all $s' \geq s$ by

$$\tilde{u}_{\pi^{-1}(s'),S-s+2}(j) \triangleq c_{j,\pi^{-1}(s')}, \quad j \in [K_{s-1,s}].$$

4. The vectors $\mathbf{u}_{\pi^{-1}(s'),S-s+2}$ are determined for all $s' \geq s$ by the binary representation of $\tilde{\mathbf{u}}_{\pi^{-1}(s'),S-s+2}$.

Based on successive cancellation at the current decoding stage, the k_s bits corresponding to the information set $\mathcal{A}_n^{(s)}$ are decoded. This completes the decoding of all the binary vectors $\mathbf{u}_{\pi^{-1}(s),l}$ for $l \in [S]$.

Remark 6.7 (On channels with equal capacities) The case where for an index $s' \in [S]$, $C_{s'} = C_{s'+1}$ is treated by skipping the construction of $\mathcal{C}_{s'}$. The coset codewords are defined by

$$\begin{aligned}\mathbf{x}_s = &\sum_{l=1}^{s'-1} \mathbf{u}_{s,l} G_n \left(\mathcal{A}_n^{S-(l-1)} \setminus \mathcal{A}_n^{S-(l-2)}\right) \\ &+ \mathbf{u}_{s,s'+1} G_n \left(\mathcal{A}_n^{(S-s')} \setminus \mathcal{A}_n^{(S-s'+2)}\right) \\ &+ \sum_{l=s'+2}^{S} \mathbf{u}_{s,l} G_n \left(\mathcal{A}_n^{(S-(l-1))} \setminus \mathcal{A}_n^{(S-(l-2))}\right) + \mathbf{b} G_n \left([n] \setminus \mathcal{A}_n^{(1)}\right), \quad s \in [S]\end{aligned}$$

At the decoding stage, two consecutive polar successive cancellation decoding can be performed for both vectors received at the output of the channel $P_{s'}$ and $P_{s'+1}$.

6.3.4 A Capacity-approaching property

Theorem 6.2 The provided parallel coding scheme achieves the capacity of every arbitrarily-permuted memoryless degraded and symmetric set of parallel channels.

Proof: Consider a set of S arbitrary-permuted degraded memoryless parallel channels P_s, $s \in [S]$, whose capacities are C_s, $s \in [S]$, respectively, and assume that the channels are ordered so that

$$C_1 \geq C_2 \geq \cdots \geq C_S.$$

According to Theorem 6.1, the capacity C_{Π} for the considered model is equal to the sum in (6.3). For a rate $R < C_{\Pi}$, choose a rate set $\{R_s\}_{s=1}^{S}$ satisfying

$$\begin{aligned} R_s &< C_s \\ \sum_{s=1}^{S} R_s &\geq R. \end{aligned} \tag{6.20}$$

The parallel polar coding in Section 6.3.3 is considered. The rate of the proposed scheme is given by

$$\frac{1}{n} \sum_{s=1}^{S} |\mathcal{A}_n^{(s)}|.$$

From (6.7) and (6.20), it follows that the proposed scheme can be designed to operate at every rate below capacity. It is left to prove that the block error probability of the proposed scheme can be made arbitrarily small for a sufficiently large block length.

Consider the vectors

$$\mathbf{u}_{s,l}, \quad s, l \in [S] \tag{6.21}$$

in (6.18). These vectors include all the information bits to be transmitted (in addition to coded versions of these bits). These vectors are determined either via the successive cancellation decoding procedure of the polar codes, or determined by the MDS code structure applied in the parallel scheme. The successive cancellation decoding procedure is based on detecting the input to the set of split channels $P_{s,n}^{(l)}$ where $s \in [S]$ and $l \in \mathcal{A}_n^{(s)}$. The information bit corresponding to a split channel $P_{s,n}^{(l)}$, is denoted by $a_{s,l}$. Note that the bit $a_{s,l}$ is either determined by the successive cancellation decoding procedure for polar codes, or else determined by the codeword of an MDS code for which it belongs to. In cases where the bit $a_{s,l}$ is decoded via a polar successive cancellation decoding procedure, the decoded bit is denoted by $\hat{a}_{s,l}$.

The bits decoded via polar successive cancellation decoding procedure, based on the received vector at the output of the channel P_s, $s \in [S]$, are

$$\hat{a}_{s,l}, \quad l \in \mathcal{A}_n^{(s)}. \tag{6.22}$$

Note that the bits in (6.22) do not include all the bits in (6.21). Nevertheless, the rest of the bits in (6.21) are fully determined from the decoded bits in (6.22) based on the MDS code structure (as detailed in the previous section).

Assuming that a permutation π is applied to the transmission of codewords, define the events

$$\mathcal{F}_{s,l} \triangleq \left\{ \hat{a}_{s,l} \neq a_{\pi^{-1}(s),l},\ \hat{a}_{s',l'} = a_{\pi^{-1}(s'),l'} : \quad \text{for all } s' \leq s, l' < l \right\}$$

where $s \in [S]$ and $l \in \mathcal{A}_n^{(s)}$. Since all the information bits can be fully determined from the bits in (6.22), the conditional block error probability is given by

$$P_{\mathrm{e}|m} = \Pr\left(\cup_{s=1}^{S} \cup_{l \in \mathcal{A}_n^{(s)}} \mathcal{F}_{s,l} \right)$$

where m is the transmitted message (representing the k information bits). According to Proposition 5.3, the events $\mathcal{E}_l(P_s)$ for $s \in [S]$ and $l \in \mathcal{A}_n^{(s)}$, defined in (5.16), are independent of the transmitted message. Moreover, it follows that

$$\mathcal{F}_{s,l} \subseteq \mathcal{E}_l(P_s).$$

Consequently, the average block error probability is upper bounded using the union bound according to

$$P_{\mathrm{e}} \leq \sum_{s \in [S]} \sum_{l \in \mathcal{A}_n^{(s)}} \Pr\left(\mathcal{E}_l(P_s)\right) \tag{6.23}$$

Finally, plugging the upper bound on the error probability (6.9) into (6.23), assures that for every fixed $S > 0$, the block error probability can be made arbitrarily low as the block length increases. ■

6.4 Parallel Polar Coding for Non-Degraded Parallel Channels

6.4.1 Signaling over Parallel Erasure Channels

The following proposition, provided in [56], considers the Bhattacharyya parameters of the split channels:

Proposition 6.2 (On the worst Bhattacharyya parameter) [56] Let p be a binary-input memoryless output-symmetric channel, and consider the split channel $p_n^{(l)}$ where $l \in [n]$. Then, among all such binary-input memoryless output-symmetric channels p whose Bhattacharyya parameter equals B, the binary erasure channel has the maximal Bhattacharyya parameter $B(p_n^{(l)})$, for every $l \in [n]$.

The proof of Proposition 6.2 is based on a tree-channel characterization of split channels, in addition to an argument which is related to extremes of information

combining. Based on Proposition 6.2, a polar signaling scheme is provided in [56] for reliable communication in a compound setting. A similar technique is used in the following for the parallel channel setting.

Consider the parallel transmission model in Section 6.1.1. In this section, it is assumed that the parallel channels are binary-input memoryless and symmetric, but are not necessarily degraded. We further assume, without loss of generality, that the set of parallel channels $\{P_s\}_{s\in[S]}$, are ordered such that

$$B(P_1) \leq B(P_2) \leq \ldots \leq B(P_S)$$

where $B(P_s)$ is the Bhattacharayya parameter of the channel P_s, $s \in [S]$ (note that the Bhattacharyya parameter varies from 0 to 1 with extremes of zero and one for a noiseless and completely noisy channels, respectively). Next, consider the set of parallel binary erasure channels, $\{\delta_s\}_{s\in[S]}$ where the erasure probability of the channel δ_s equals $B(P_s)$, $s \in [S]$. These erasure channels form a family of S stochastically degraded channels. Consequently, based on Theorem 6.2, the parallel polar coding scheme in Section 6.3.3 achieves a rate of $S - \sum_{s=1}^{S} B(P_s)$ over the set of erasure channels, under the successive cancellation decoding scheme detailed in Section 6.3.3. The following corollary addresses the performance of the same coding scheme over the original set of parallel channels:

Corollary 6.3 The polar coding scheme for the parallel erasure channels, operates reliably over the original parallel channels.

Proof: The suggested coding scheme performs reliably over the parallel binary erasure channels. The decoding process, as described in Section 6.3.3, includes a sequence of successive cancellation decoding operations applied to the polar codes over each one of the parallel channels. As shown in the proof of Theorem 6.2, reliable communication is obtained based on reliably decoding each of the successive cancellation operations. It is therefore required to show that the successive cancellation over the original channels $\{P_s\}_{s\in[S]}$ can also be carried reliably, this follows as a consequence of Proposition 6.2. Denote the sequences of information sets chosen for reliable communication over the erasure channels $\{\delta_s\}_{s\in[S]}$ by $\{\mathcal{A}_n^{(s)}\}_{s\in[S]}$. Each one of these sets satisfies the properties in Theorem 5.3. Fix an arbitrary channel P_s from the set of parallel channels, and an arbitrary index $l \in A_n^{(s)}$. Consider next the error event $\mathcal{E}_l(P_s)$ in (5.16). According to Proposition 5.3, this error event is upper bounded by

$$\Pr\big(\mathcal{E}_l(P_s)\big) \leq B\big((P_s)_n^{(l)}\big) \tag{6.24}$$

where $B\big((P_s)_n^{(l)}\big)$ denotes the Bhattacharayya parameter of the split channel $(P_s)_n^{(l)}$. From Proposition 6.2, it follows that

$$B\big((P_s)_n^{(l)}\big) \leq B\big((\delta_s)_n^{(l)}\big) \tag{6.25}$$

where $B\big((\delta_s)_n^{(l)}\big)$ is the Bhattacharayya constant of the split channel $(\delta_s)_n^{(l)}$. Fix $0 < \beta < \frac{1}{2}$ as in Theorem 5.3. From (6.24) and (6.25), it follows from Theorem 5.3 that

$$\Pr\big(\mathcal{E}_l(P_s)\big) \leq 2^{-n^\beta}.$$

Consequently, the successive cancellation decoding operations can be carried reliably for each one of the original channels, which completes the proof. ■

6.4.2 A Compound Interpretation of Monotone Index Set Design and Related Results

The parallel coding scheme provided in Section 6.3 is based on a monotonic sequence of index sets $\{\mathcal{A}_n^{(s)}\}_{s\in[S]}$ satisfying the conditions in Corollary 6.2. As explained in Remark 6.5, the index sets in $\mathcal{A}_n^{(s)}$, $s \in [S]$ are 'good' for all the channels $P_{s'}$, $s' \geq s$. Here, as in Remark 6.5, 'good' means that the corresponding Bhattacharayya parameters of the corresponding split channels satisfy the polarization properties in Theorem 5.3. The index set sequences $\{\mathcal{A}_n^{(s)}\}_{s\in[S]}$ are applied in this paper to parallel transmission. Even though the compound setting and the problem of parallel transmissions are at first glance different, the actual problem of finding an index sets which is 'good' for a set of channels is similar to the problem studied in [56] in the compound model.

In the compound setting, the transmission takes place over one channel which belongs to a predetermined set of channels. It is assumed in the current discussion that (only) the receiver knows the channel over which the transmission takes place. If a polar code is applied in such a compound setting, then a suitable index set is required. Such an index set must be 'good' for all the channels in the set. The maximal rate over which such a polar coding scheme performs reliably is termed as the compound capacity of polar codes. Obviously, the compound capacity relates to the size of possible 'good' index sets.

Upper and lower bounds on the compound capacity of polar codes under successive cancelation decoding are provided in [56]. These bounds are defined using the notion of tree-channels. Let p be a binary-input memoryless output-symmetric channel. For a binary vector of length k, $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_k)$, the tree-channel associated to σ is denoted by p^σ. The actual definition of the tree-channel is not required for

the following discussion, and is therefore omitted (the reader is referred to [56] and references therein for more details). It is noted that the tree-channel is also binary-input memoryless and output-symmetric. Moreover, it is further noted in [56] that the tree-channel p^{σ}, is equivalent to the split-channel $p_n^{(l)}$ where σ is the binary expansion of l.

Let $\{P_s\}_{s\in[S]}$ be a set of S binary-input memoryless output-symmetric channels. It is shown in [56] that the compound capacity for the considered setting $C\big(\{P_s\}_{s\in[S]}\big)$ is lower bounded by[1]

$$C\big(\{P_s\}_{s\in[S]}\big) \geq 1 - \frac{1}{2^k} \sum_{\sigma\in\{0,1\}^k} \max_{s\in[S]} B\big(P_s^{\sigma}\big) \tag{6.26}$$

where $k \in \mathbb{N}$ and $B\big(P_s^{\sigma}\big)$ is the Bhattacharyya parameter of the tree-channel P_s^{σ}. Moreover, this lower bound is a constructive bound. That is, the construction of an appropriate index set sequence $\mathcal{A}_n\big(\{P_s\}_{s\in[S]}\big)$ is inherent from the lower bound. The polar code corresponding to this index set has an asymptotically low decoding error probability under successive cancellation decoding (for every channel in the set $\{P_s\}_{s\in[S]}$).

Corollary 6.4 (**Improved parallel polar coding scheme**) Consider the transmission over a set of parallel binary-input memoryless and output-symmetric channels $\{P_s\}_{s\in[S]}$. Fix an order $P_{s_1}, P_{s_2}, \ldots, P_{s_S}$ of channels and $k \in \mathbb{N}$. Then, reliable transmission is achievable based on the parallel polar coding scheme in Section 6.3, whose rate is given by

$$C(P_{s_S}) + S - 1 - \frac{1}{2^k} \sum_{s\in[S-1]} \sum_{\sigma\in\{0,1\}^k} \max_{i\in\{s,\ldots S\}} B\big(P_{s_i}^{\sigma}\big). \tag{6.27}$$

Proof: Define the channel sets

$$\mathcal{P}_s \triangleq \{P_{s_i}\}_{i=s}^{S}, \quad s \in [S].$$

[1]The actual derivation in [56] is provided for two channels P and Q. Nevertheless, the arguments in [56] are suitable for the case of $S > 2$ channels. The proof of the bounds in [56] is based on two major arguments. The first argument consider a sequential transformations of a given channel P to a sequence of sets of tree-channels. Initially, the channel P is transformed into a pair of tree-channels P^0 and P^1. Next, each of these tree-channels is transformed again to another pair, and the transformation repeats recursively. It is shown that instead of transmitting bits corresponding to indices induced by the polarization of the original channel P, at each transformation level k, the problem is equivalent to transmitting a fraction $\frac{1}{2^k}$ of the bits based on the indices induced by the polarization of the corresponding tree channels $\{P^{\sigma}\}_{\sigma\in\{0,1\}^k}$. The first argument is therefore not affected by the number of channels (as it concerns a property of a single channel). The second argument is identical to the more simple polarization scheme detailed in Section 6.4.1. This polarization scheme, based on binary erasure channels, can be applied to every set of tree-channels $\{P_s^{\sigma}\}_{s=1}^{S}$, $\sigma \in \{0,1\}^k$. Based on this polarization scheme, a rate of $\frac{1}{2^k}\left(1 - \max_{s\in[S]} B\big(P_s^{\sigma}\big)\right)$ is guaranteed for each $\sigma \in \{0,1\}^k$.

For each channel set $\mathcal{P}_s$, $s \in [S]$, the compound setting is considered. Based on the lower bound in (6.26) and its associated index set sequence, a set sequence $\mathcal{A}_n(\mathcal{P}_s)$ exists for every $s \in [S]$, such that

$$\frac{1}{n}\mathcal{A}_n(\mathcal{P}_s) \geq 1 - \frac{1}{2^k} \sum_{\sigma \in \{0,1\}^k} \max_{i \in \{s,\dots S\}} B\big(P_{s_i}^{\sigma}\big) \tag{6.28}$$

and reliable decoding is guaranteed for all the channels in the set $\mathcal{P}_s$ under successive cancellation decoding. As an immediate consequence of the construction, for every n, the index sets form a monotonic sequence (i.e., if an index is 'good' for a set of channels, it must be 'good' for a subset of these channels). Therefore, the monotone set sequences for the polar construction is provided and the parallel polar scheme in Section 6.3 can be applied. The rate of the resulting scheme is given by summing over the rates in (6.28) which adds to

$$S - \frac{1}{2^k} \sum_{s \in [S]} \sum_{\sigma \in \{0,1\}^k} \max_{i \in \{s,\dots S\}} B\big(P_{s_i}^{\sigma}\big).$$

Since the last channel set $\mathcal{P}_S$ includes just a single channel P_{s_S}, the compound setting is not required for this set. For the last set the information index set of the polar coding construction (in Section 5.1.2) is therefore applied. The resulting rate of the parallel scheme is improved and given by (6.27). ∎

Remark 6.8 (**Possible order of channels**) The channel order may be an important parameter for the provided parallel scheme (in terms of achievable rates). The channels may be ordered by their capacity, where

$$C(P_{s_1}) \leq C(P_{s_2}) \leq \cdots \leq C(P_{s_S}).$$

However, we have no evidence that this order results in the maximal achievable rate (or that it is optimal in any other sense).

Remark 6.9 (**An upper bound on parallel polar capacity**) For each set $\mathcal{P}_s$, $s \in [S]$, the upper bound in [56] on the compound capacity can be applied to upper bound the size of the existing index sets $\mathcal{A}_n(\mathcal{P}_s)$. According to [56, Theorem 5], the resulting rate is upper bounded by[2]

$$\frac{1}{2^k} \sum_{\sigma \in \{0,1\}^k} \min_{i \in \{s,\dots,S\}} I\big(P_{s_i}^{\sigma}\big)$$

[2]As in the case of the lower bound, the actual derivation in [56] is provided for two channels P and Q. Nevertheless, the arguments in [56] are suitable for the case of $S > 2$ channels. The proof of the considered upper bound is based on two major arguments. The first argument is a transformation of a channel to a sequence of sets of tree-channels (the same as in the lower bound). Then, for each such set, the maximal achievable rate is upper bounded by the minimal capacity of the channel capacities.

for every $k \in \mathbb{N}$, where $I\big(P_{s_i}^{\sigma}\big)$ is the capacity of the corresponding tree-channel $P_{s_i}^{\sigma}$. Since for the last channel set, which is a set of a single channel, we have no compound setting (as explained in the proof of Corollary 6.4) the maximal rate at which the parallel polar coding scheme proposed in Section 6.3 can operate reliably is given by

$$C(P_{s_S}) + \frac{1}{2^k} \sum_{s \in [S-1]} \sum_{\sigma \in \{0,1\}^k} \min_{i \in \{s, \dots S\}} I\big(P_{s_i}^{\sigma}\big).$$

An example is provided in [56], demonstrating the the concerned bound can be smaller than each of the channel capacities. Specifically, the example in [56] is based on a BSC with a crossover probability of 0.11002 and a BEC whose erasure probability is 0.5. Both of these channels corresponds to a capacity of 0.5 bits per channel use. However, as demonstrated in [56, Example 6], their compound capacity is upper bounded by 0.482 bits per channel use. Consequently, if the parallel polar coding scheme in Section 6.3 is applied for the same two channels, the possible rate of such a parallel coding scheme is upper bounded by 0.982 bits per channel use where the parallel capacity is given by 1 bit per channel use.

6.5 Summery and Conclusions

A parallel polar coding scheme is provided in this chapter for binary-input arbitrarily-permuted memoryless and output-symmetric parallel channels. The provided polar codes are shown to achieve the capacity of the considered model where the channels are assume to be stochastically degraded. For the non-degraded case an upper and lower bounds on the achievable rates are provided. A generalization to non-binary parallel polar coding, based on the results in [91], is clear.

Chapter 7

Summary and Outlook

7.1 Summary

The performance of non-binary linear block codes under ML decoding is analyzed in Chapter 2. We provided a definition of symmetry for memoryless channels with non-binary input alphabets. Under the provided symmetry condition, we proved that the conditional error probability under ML decoding is independent of the transmitted codeword. This result generalizes the well known message-independence property for MBIOS channels (see also [39] and [40] where the same result was proved under linear-programming decoding). The main part of Chapter 2 is devoted to the derivation of upper bounds on the error performance of linear block codes under ML decoding. We next apply these bounds on ensembles of regular non-binary LDPC codes, and study their error performance for various communication channel models. In addition, we provide the exact complete composition spectra for these LDPC code ensembles (instead of the upper bound in [44]). This analysis forms a generalization of [18] and [105] in the binary setting. Finally, we compare the new upper bounds with sphere-packing lower bounds on the decoding error probability, and show that the bounds are informative even at the low SNR regime.

In Chapter 3 we provide upper bounds on the error probabilities under generalized decoding rules, i.e., list decoding rules and decoding rules with erasures. Our bounds are valid for linear block codes whose transmission takes place over memoryless symmetric channels. We also provide message independence results for the considered generalized decoding rules where both optimal and suboptimal decoding rules are considered. When variable-size list-decoding is considered, we derive upper bounds on the expected size of the decoded list and the associated error probability under list decoding. In addition, upper bounds on the list error probability of linear block codes are introduced when the size of the list is fixed. The bounds derived in this chapter

are applicable to the performance analysis of specific codes and code ensembles, via their (average) distance spectra. The bounds are suitable for finite block lengths and also for asymptotic analysis. We finally exemplify the bounds for two coding schemes: Fully-random linear block codes, and regular (binary and non-binary) LDPC code ensembles with finite block lengths. We also exemplify the applications of the bounds to hybrid-ARQ schemes.

In Chapter 4 we study possible generalizations of the VA for list-decoding and decoding with erasures. We introduce a modification of the VA, which coincides with the optimal decoding rule of Forney for the cases at hand. The new algorithm applies to the more general case where finite-state Markov processes are observed via memoryless channels, while our presentation is focused on the decoding of convolutional codes. We simulated the performance of the proposed modified algorithm and compared the results with the simulated performance of two suboptimal decoding algorithm with erasures: the likelihood-ratio (LR) test decoding rule, and a simple decoding scheme with repeat requests provided by Yamamoto and Itoh in [120]. A good similarity between the performance of the simple scheme to the optimal one was observed, even though the decoding scheme in [120] is remarkably simple. On the other hand, the performance of the decoding algorithm based on the LR test is found to be considerably degraded in comparison with that of the optimal performance.

In Chapter 5, we study the application of channel polarization to the wire-tap communication model. We show that the secrecy capacity of a degraded memoryless binary-input and symmetric wire-tap channel can be achieved by a proper application of the channel polarization method. We prove that for every rate below the channel secrecy capacity, there exists a suitable polar code for which both conditions of reliable and secure communication are achieved under successive cancelation decoding. In addition, we prove that the entire equivocation-rate can be achieved with channel polarization under the weak notion of secrecy. For the particular case of erasure wire-tap channel, with perfect observations at the legitimate user, it is shown that the secrecy capacity can be achieved with a strong notion of secrecy. Finally, we study the possible application of channel polarization for non-degraded wire-tap channels.

In Chapter 6, we continue the study of some possible applications of channel polarization, by considering the signaling over parallel channels. We propose a channel coding scheme and its corresponding successive cancelation decoding algorithm for signaling over parallel channels. In the proposed scheme, the method of channel polarization is incorporated with an algebraic maximum-distance separable codes. In addition, it is shown that by using the proposed coding scheme, the capacity of signaling over arbitrarily-permuted memoryless and symmetric parallel-channels is

achievable under an assumption of channel degradation. Finally, the assumption on channel degradation is excluded.

7.2 Outlook

Performance bounds of non-binary coding schemes

The comparison of the bounds in Chapter 2 with sphere-packing lower bounds shows the suitability of the new bounds for the study of capacity approaching non-binary linear block codes whose transmission takes place over an AWGN channel. For the case of fully-interleaved fading channels with perfect CSI at the receiver, the comparison of these upper and lower bounds shows a gap which motivates further study of analysis techniques for non-binary linear block codes. It is unclear if the observed gap is due to the provided bounding technique or is it a question of the suitability of the specific codes which may not be the best choice for the fully interleaved fading channel model. Hence, the comparison of additional non-binary coding techniques with the provided bounding technique is of interest. In particular, the adaptation and possible generalization for the definition of symmetry and the related message-independence property for further non-binary modulation techniques is of great interest.

Variations on recently introduced random-coding bounds

The recently introduced coding theorems in [84] and [114] enable to derive improved upper bounds on the error performance of coded schemes. In [84], new performance bounds are derived for general channels (both achievable and converse results are provided). These bounds are tighter than classical bounds and some recently introduced bounds. The achievable results in [84] and [114] are derived via the random coding technique. The bounds derived in this thesis are based on some variations of the Gallager bounding technique which was originally derived for random code ensembles. It is of interest to adapt the recently introduced technique such that it may serve for the performance analysis of structured codes.

Generalized decoding of linear block codes

The analysis of coded communications over fully-interleaved fading channels with CSI at the receiver under generalized decoding is of interest. Moreover, the suitability of the provided bounds for such communication channel models is apparat. The analysis of the error and latency performance of coded communication systems over fading

channels with feedback, based on the provided bounds, is of special interest. Possible adaptation of the bounds for the case of transmissions over parallel channels is also of interest.

Providing feasible decoding algorithms with variable list sizes and erasures is of major interest. The literature on list decoding and decoding with ARQ schemes is tremendous. However, it seems that much less work is devoted to the analysis and design of decoding algorithms which aim to operate close to the optimal decoding rule in [41]. Moreover, we still lack a good understanding of efficient, practically appealing coding and decoding schemas in the realm of different degrees of feedback (in terms of feedback rate and reliability). The study of such coding scheme is of special interest for systems which do not exhibit exponential behavior of the error probability, such as LDPC under iterative decoding, but may acquire exponential behavior in the presence of feedback.

Channel polarization

The study of possible application of channel polarization for non-degraded wire-tap channels is of major interest. In particular, it is suggested to investigate the open polarization problem in Chapter 5. Generalizing the proposed applications to non-binary and non-symmetric channels, via the non-binary channel polarization method in [79], [91] is also suggested as a continuation of major interest. Variations on the parallel polar coding technique which may achieve the capacity of some non-degraded parallel channels are of interest. The way channel polarization is combined with algebraic codes for parallel channels is a technique which may contribute to some further applications, besides the particular case of parallel channels (see, e.g., [80]). This technique may also help in improving the error exponents of the channel polarization method, and in particular its successive decoding algorithm.

References

[1] "Digital cellular telecommunications system (phase 2+); channel coding," GSM 05.03, ETSI EN 300 909, European Telecommunications Standards Institute 2000, version 8.5.1, 1999, 1999.

[2] A. M. A. G. Fabregas and G. Caire, *Bit-interleaved coded modulation*, ser. Foundations and Trends in Communication and Information Theory. Now Publishers, 2008, vol. 5, no. 1–2.

[3] M. Andersson, V. Rathi, R. Thobaben, J. Kliewer, and M. Skoglund, "Nested Polar Codes for Wiretap and Relay Channels," to apear in *IEEE Communications Letters*.

[4] E. Arikan, "Channel polariazation: a method for constructing capacity-acheiving codes for symmetric binary-input memoryless channels," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3051–3073, July 2009.

[5] E. Arikan and G. Markarian, "Two dimensional polar coding," in *Proceedings 2009 International Sympossium on Communication Theory and Applications (ISCTA 2009)*, Ambleside, UK, July 2009.

[6] E. Arikan and E. Telatar, "On the rate of channel polarization," in *Proceedings 2009 IEEE International Symposium on Information Theory (ISIT 2009)*, Seol, Korea, June 2009.

[7] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Transactions on Information Theory*, vol. 20, no. 2, pp. 284–287, March 1974.

[8] C. Bai, B. Mielczarek, W. Krzymien, and I. Fair, "Improved analysis of list decoding and its application to convolutional codes and turbo codes," *IEEE Transactions on Information Theory*, vol. 53, no. 2, pp. 615–627, February 2007.

[9] A. Barg, "Improved error bounds for the erasure/list scheme: The binary and spherical cases," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2503–2511, October 2004.

[10] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Bandwidth efficient parallel concatenated coding schemes," *Electronics Letters*, vol. 31, no. 24, pp. 2067–2069, 1995.

[11] 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.

[12] ——, "Design and analysis of nonbinary ldpc codes for arbitrary discrete memoryless channels," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 549–583, February 2006.

[13] 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.

[14] E. Biglieri, D. Divsalar, M. Simon, P. McLane, and J. Griffin, *Introduction to trellis-coded modulation with applications.* Prentice-Hall, Inc. Upper Saddle River, NJ, USA, 1991.

[15] R. Blahut, *Algebraic codes for data transmission.* Cambridge University Press, 2003.

[16] E. L. Blokh and V. V. Zyablov, *Linear Concatenated Codes.* (in Russian). Moscow, U.S.S.R.: Nauka, 1982.

[17] I. Bocharova, R. Johannesson, B. Kudryashov, and M. Loncar, "An improved bound on the list error probability and list distance properties," *IEEE Transactions on Information Theory*, vol. 54, no. 1, pp. 13–32, January 2008.

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

[19] G. Caire, S. Shamai, and S. Verdú, "Feedback and belief propagation," in *Proceedings of the 2006 Turbo Coding Symposium*, Munich, Germany, April 2006, pp. 3–7.

[20] B. Chen and C. Sundberg, "List Viterbi algorithms for continuous transmission," *IEEE Transactions on Communications*, vol. 49, no. 5, pp. 784–792, May 2001.

[21] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*. New York: Springer, 1988.

[22] D. J. Costello and G. D. Forney, "Channel coding: The road to channel capacity," *Proceedings of the IEEE*, vol. 95, no. 6, pp. 1150–1177, June 2007.

[23] T. Cover and J. Thomas, *Elements of information theory*. Wiley, 2006.

[24] I. Csiszar and J. Korner, *Information theory: coding theorems for discrete memoryless systems*. ANew York: Academic, 1981.

[25] M. Davey and D. Mackay, "Low-density parity check codes over $gf(q)$," *IEEE Communications Letters*, vol. 2, no. 6, pp. 165–167, June 1998.

[26] D. Divsalar, "A simple tight bound on error probability of block codes with application to turbo codes," Jet Propulsion Laboratory (JPL), the Telecommunications and Mission Operations (TMO) Progress Report 42-139, November 1999.

[27] D. Divsalar and E. Biglieri, "Upper bounds to error probabilities of coded systems beyond the cutoff rate," *IEEE Transactions on Communications*, vol. 51, no. 12, pp. 2011–2018, December 2003.

[28] S. Dolinar, K. Andrews, F. Pollara, and D. Divsalar, "Bounds on error probability of block codes with bounded-angle maximum-likelihood incomplete decoding," in *Proceedings 2008 Interntional Symposium on Information Theory and its Applications*, Auckland, New Zeland, December 2008.

[29] ——, "The limits of coding with joint constraints on detected and undetected error rates," in *Proceedings 2008 IEEE International Symposium on Information Theory (ISIT 2008)*, Toronto, Canada, July 2008, pp. 970–974.

[30] S. Draper and A. Sahai, "Variable-length coding with noisy feedback," *European Transactions on Telecommunications*, vol. 19, no. 4, pp. 355–370, April 2008.

[31] T. M. Duman, "Turbo codes and turbo-coded modulation systems: Analysis and performance bounds," Ph.D. dissertation, Northeastern University, Boston, MA, May 1998.

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

[33] ——, "Performance bounds for turbo-coded modulation systems," *IEEE Transactions on Communications*, vol. 47, no. 4, pp. 511–521, April 1999.

[34] I. S. E. Hof and S. S. (Shitz), "Performance bounds for non-binary linear block codes over memoryless symmetric channels," *IEEE Transactions on Information Theory*, vol. 55, no. 3, pp. 977–996, March 2009.

[35] P. Elias, "Coding for noisy channels," *IRE Conv*, vol. 4, pp. 37–46, March 1955.

[36] ——, "List decoding for noisy channels," Research Laboratory of Electronics, Massachusetts Institute of Technology (MIT), Tech. Rep. 335, September 1957.

[37] U. Erez and G. Miller, "The ml decoding performance of ldpc ensembles over $\mathbb{Z}_q$," *IEEE Transactions on Information Theory*, vol. 51, no. 5, pp. 1871–1879, May 2005.

[38] R. Fano, "A heuristic discussion of probabilistic decoding," *IEEE Transactions on Information Theory*, vol. 9, no. 2, pp. 64–74, January 1963.

[39] M. Flanagan, "Codeword-independent performance of nonbinary linear codes under linear-programming and sum-product decoding," in *Proceedings 2008 IEEE International Symposium on Information Theory (ISIT 08)*, Toronto, Canada, July 2008, pp. 1503–1507.

[40] M. Flanagan, V. Skachek, E. Byrne, and M. Greferath, "Linear-programming decoding of non-binary linear codes," in *Proceedings 7th International Conference on Source and Channel Coding (SCC2008)*, Ulm, Germany, January 2008, pp. 14–16.

[41] G. Forney, "Exponential error bounds for erasure, list, and decision feedback schemes," *IEEE Transactions on Information Theory*, vol. 14, no. 2, pp. 206–220, March 1968.

[42] ——, "The viterbi algorithm," *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268–278, March 1973.

[43] J. G. D. Forney, *Concatenated Codes*. Cambridge, MA: MIT Press, 1966.

[44] R. G. Gallager, "Low-density parity-check codes," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA, 1963.

[45] ——, "A simple derivation of the coding theorem and some applications," *IEEE Transactions on Information Theory*, vol. 11, no. 1, pp. 3–18, January 1965.

[46] ——, *Information Theory and Reliable Communications*. John Wiley and Sons, 1968.

[47] M. J. E. Golay, "Notes on digital coding," *Proceedings of the IRE*, vol. 37, p. 657, January 1949.

[48] P. K. Gopala, N. Young-Han, and H. El-Gamal, "On the error exponents of ARQ channels with deadlines," *IEEE Transactions on Information Theory*, vol. 53, no. 11, pp. 4265–4273, November 2007.

[49] R. W. Hamming, "Error detecting and error correcting codes," *Bell System Technical Journal*, vol. 29, no. 2, pp. 147–160, April 1950.

[50] T. Hashimoto, "A list-type reduced-constraint generalization of the Viterbi algorithm," *IEEE transactions on information theory*, vol. 33, no. 6, pp. 866–876, September 1987.

[51] ——, "A coded ARQ scheme with the generalized Viterbi algorithm," *IEEE Transactions on Information Theory*, vol. 39, no. 2, pp. 423–432, March 1993.

[52] ——, "Comparison of erasure-and-error threshold decoding schemes," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 76, no. 5, pp. 820–827, May 1993.

[53] ——, "On the error exponent of convolutionally coded ARQ," *IEEE Transactions on Information Theory*, vol. 40, no. 2, pp. 567–575, March 1994.

[54] ——, "Composite shceme LR + Th for decoding with erasures and its effective equivalence to Forney's rule," *IEEE Transactions on Information Theory*, vol. 45, no. 1, pp. 78–93, January 1999.

[55] S. H. Hassani, S. B. Korada, and R. Urbanke, "The Compound Capacity of Polar Codes," in *Proceedings Allerton Conference on Communication, Control and Computing,*, Monticello, IL, September 2009.

[56] S. H. Hassani, S. B. Korada, and R. Urbnake, "Nested Polar Codes for Wiretap and Relay Channels," in *Allerton Conference on Communication, Control and Computing*, Allerton, September 2009.

[57] H. Herzberg and G. Poltyrev, "Techniques of bounding the probability of decoding error for block coded modulation structures," *IEEE Transactions on Information Theory*, vol. 40, no. 3, pp. 903–911, May 1994.

[58] S. S. I. Sason and D. Divsalar, "Tight exponential upper bounds on the ml decoding error probability of block codes over fully-interleaved fading channels," *IEEE Transactions on Communications*, vol. 51, no. 8, pp. 1296–1305, August 2003.

[59] H. Imai and S. Hirakawa, "A new multilevel coding method using error-correcting codes," *IEEE Transactions on Information Theory*, vol. 23, no. 4, pp. 371–377, May 1997.

[60] R. Johannesson and K. Zigangirov, *Fundamentals of convolutional coding*. Orient Blackswan, 1905.

[61] D. Knuth, *The Art of Computer Programming. Vol. 2: Seminumerical Algorithms, (3rd edition)*. Addison-Wesley Pub. Co, 1998.

[62] S. B. Korada, "Polar codes for channel and source coding," Ph.D. dissertation, EPFL, Lausanne,Switzerland,, 2009.

[63] S. B. Korada and E. Sasoglu, "A class of transformations that polarize binary-input memoryless channels," in *Proceedings 2009 IEEE International Symposium on Information Theory (ISIT 2009)*, Seol, Korea, June 2009.

[64] S. B. Korada, E. Sasoglu, and R. Urbanke, "Polar codes: characterization of exponent, bounds, and constructions," in *Proceedings 2009 IEEE International Symposium on Information Theory (ISIT 2009)*, Seol, Korea, June 2009.

[65] S. B. Korada and R. Urbanke, "Polar Codes are Optimal for Lossy Source Coding," *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1751–1768, April 2010.

[66] O. Koyluoglu and H. El-Gamal, "Polar Coding for Secure Transmission and Key Agreement," in *to be presented at Proceedings 2010 IEEE Information Theory Workshop (ITW 2010)*, Dublin, Ireland, October 2010.

[67] B. D. Kudryashov, "Error probability for repeat request systems with convolutional codes," *IEEE Transactions on Information Theory*, vol. 39, no. 5, pp. 1680–1684, March 1993.

[68] B. Kudryashov, "List decoding in a Gaussian channel," *Problemy Peredachi Informatsii*, vol. 27, no. 3, pp. 30–38, July-September 1991.

[69] Y. Liang, H. V. Poor, and S. Shamai, *Information Theoretic Security*, ser. Foundations and Trends in Communication and Information Theory. Now Publishers, 2008, vol. 5, no. 4–5.

[70] L. Lijofi, D. Cooper, and B. Canpolat, "A reduced complexity list single-wrong-turn (SWT) Viterbi decoding algorithm," in *Proceedings of the 15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 2004 (PIMRC 2004)*, September 2004, pp. 274–279.

[71] S. Lin and D. J. Costello, *Error Control Coding*, 2nd ed. Prentice Hall, 2004.

[72] R. Liu, J. Luo, and P. Spasojevic, "Adaptive transmission with variable-rate turbo bit-interleaved coded modulation," *IEEE Transactions on Wireless Communications*, vol. 6, no. 11, pp. 3926–3936, 2007.

[73] 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.

[74] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Amsterdam, The Netherlands: North Holland, 1977.

[75] H. Mahdavifar and A. Vardy, "Achieving the Secrecy Capacity of Wiretap Channels Using Polar Codes," in *to be presented at Proceedings 2010 IEEE International Symposium on Information Theory (ISIT 2010)*, Austin, Texas, June 2010.

[76] J. L. Massey, *Threshold Decoding*. Cambridge, MA: MIT Press, 1963.

[77] N. Merhav, "Error exponents of erasure/list decoding revisted via moments of distance enumerators," *IEEE Transactions on Information Theory*, vol. 54, no. 10, pp. 4439–4447, October 2008.

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

[79] R. Mori and T. Tanaka, "Channel Polarization on q-ary Discrete Memoryless Channels by Arbitrary Kernels," in *Proceedings 2010 IEEE International Symposium on Information Theory (ISIT 2010)*, Austin, Texas,, June 2010.

[80] ——, "Non-Binary Polar Codes using Reed-Solomon Codes and Algebraic Geometry Codes," in *Proceedings 2010 IEEE Information Theory Workshop (ITW 2010)*, Dublin, Irlenad,, August-September 2010.

[81] C. Nill and C. Sundberg, "List and soft symbol output Viterbi algorithms: extensions and comparisons," *IEEE Transactions on Communications*, vol. 43, no. 2/3/4, pp. 277–287, February/March/April 1995.

[82] L. H. Ozarow and A. D. Wyner, "Wire-tap channel II," *Bell System Technical Journal*, vol. 63, December 1984.

[83] E. Perron, S. Theory, and E. Telatar, "A Multiple Access Approach for the Compound Wiretap Channel," in *Proceedings 2009 IEEE Information Theory Workshop (ITW 2009)*, Taormina, Sicily, October 2009.

[84] Y. Polyanskiy, V. Poor, and S. Verdú, "Channel Coding Rate in the Finite Blocklength Regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[85] V. Rathi and R. Urbanke, "Density evolution, stability condition, thresholds for non-binary ldpc codes," *IEEE Transactions on Information Theory*, vol. 152, no. 6, pp. 1069–1074, December 2005.

[86] T. Richardson and R. Urbanke, *Modern coding theory*. Cambridge Univ Pr, 2008.

[87] P. Robertson and T. Worz, "Bandwidth-efficient turbo trellis-coded modulation using punctured component codes," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 2, pp. 206–218, February 1998.

[88] M. Röder and R. Hamzaoui, "Fast tree-trellis list Viterbi decoding," *IEEE Transactions on Communications*, vol. 54, no. 3, pp. 453–461, March 2006.

[89] R. M. Roth, *Introduction to Coding Theory*. Cambridge University Press, 2006.

[90] W. Ryan and S. Lin, *Channel Codes: Classical and Modern*. Cambridge Univ Pr, 2009.

[91] E. Sasoglu, E. Telatar, and E. Arikan, "Polarization for arbitrary discrete memoryless channels," in *Proceedings 2009 Information Theorey Workshop (ITW 2009)*, Taormina, Sicily, Octuber 2009.

[92] 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.

[93] 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.

[94] ——, *Performance Analysis of Linear Codes under Maximum-Likelihood Decoding: A Tutorial*, ser. Foundations and Trends in Communication and Information Theory. Now Publishers, 2006, vol. 3, no. 1–2. [Online]. Available: http://www.ee.technion.ac.il/people/sason/monograph_postprint.pdf

[95] N. Seshadri and C. Sundberg, "List Viterbi decoding algorithms with applications," *IEEE Transactions on Communications*, vol. 42, no. 2/3/4, pp. 313–323, February/March/April 1994.

[96] S. Shamai and I. Sason, "Variations on the gallager bounds, connections and applications," *IEEE Transactions on Information Theory*, vol. 48, no. 12, pp. 3029–3051, December 2002.

[97] 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.

[98] C. E. Shannon, "A mathematical theory of communications," *Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, July / October 1948.

[99] ——, "Probability of error for optimal codes in a Gaussian channel," *Bell System Technical Journal*, vol. 38, no. 3, pp. 611–656, May 1959.

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

[101] B. Sklar, "How I learned to love the trellis," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 87–102, 2003.

[102] T. M. T. Niinomi and S. Hirasawa, "On the generalized Viterbi algorithm using likelihood ratio testing," in *Proceedings of the 2002 IEEE International Symposiom on Information Theory (ISIT 2002)*, Lausanne, Switzerland, July 2002.

[103] E. Telatar, "Exponential bounds for list size moments and error probability," in *Proceedings of the 1998 IEEE Information Theory Workshop (ITW 1998)*, Killarney, Ireland, June 1998, p. 60.

[104] E. Telatar and R. Gallager, "New exponential upper bounds to error and erasure probabilities," in *Proceedings 1994 IEEE International Symposium on Information Theory (ISIT 1994)*, Trondheim, Norway, June 1994, p. 379.

[105] S. Tong, "Tangential-sphere bounds on the ensemble performance of ML decoded Gallager codes via their exact ensemble distance spectrum," in *Proceedings 2008 IEEE International Conference on Communications (ICC 2008)*, Beijing, China, May 2008.

[106] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Transactions on Information Theory*, vol. 28, no. 1, pp. 55–67, January 1982.

[107] A. Valembois and M. Fossorier, "Sphere-packing bounds revisited for moderate block length," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 2998–3014, December 2004.

[108] A. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE transactions on Information Theory*, vol. 13, no. 2, pp. 260–269, April 1967.

[109] A. J. Viterbi, "Error bounds for the white Gaussian and other very noisy memoryless channels with generalzied decision regions," *IEEE Transactions on Information Theory*, vol. 15, no. 2, pp. 279–287, March 1969.

[110] A. Viterbi, "A personal history of the Viterbi algorithm," *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 120–142, July 2006.

[111] A. Viterbi and J. Omura, *Principles of digital communication and coding*. McGraw-Hill, Inc. New York, NY, USA, 1979.

[112] U. Wachsmann, R. Fischer, and J. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1361–1391, July 1999.

[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] L. Wang, R. Colbeck, and R. Renner, "Simple Channel Coding Bounds," in *Proceedings 2009 IEEE International Symposium on Information Theory (ISIT 2009)*, Seoul, Korea,, June-July 2009.

[115] G. Wiechman and I. Sason, "An improved sphere-packing bound for finite-length codes on symmetric memoryless channels," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 1962–1990, May 2008.

[116] F. M. J. Willems and A. Gorokhov, "Signaling over arbitrarily permuted parallel channels," *IEEE Transactions on Information Theory*, vol. 54, no. 3, pp. 1374–1382, March 2008.

[117] J. M. Wozencraft, "List decoding," Research Laboratory of Electronics, Massachusetts Institute of Technology (MIT), Quarterly Progress Report, January 1958.

[118] J. M. Wozencraft and B. Reiffen, *Sequential Decoding*. Cambridge, MA: MIT Press, 1961.

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

[120] H. Yamamoto and K. Itoh, "Viterbi decoding algorithm for convolutional codes with repeat request," *IEEE Transactions on Information Theory*, vol. 26, no. 5, pp. 540–547, March 1980.

[121] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Transactions on Communications*, vol. 40, no. 5, pp. 873–884, May 1992.

בעיות בתקשורת מקודדת: חסמים על ביצועים וקודים קוטביים

עירן חוף

בעיות בתקשורת מקודדת: חסמים על ביצועים וקודים קוטביים

חיבור על מחקר

לשם מילוי חלקי של הדרישות לקבלת תואר

דוקטור לפילוסופיה

עירן חוף

הוגש לסנט הטכניון — מכון טכנולוגי לישראל

אלול תש"ע חיפה ספטמבר 2010

חיבור על מחקר נעשה בהדרכת פרופ' יגאל ששון ופרופ' שלמה שמאי
בפקולטה להנדסת חשמל

הכרת תודה

ברצוני להודות למנחים פרופ' יגאל ששון ופרופ' שלמה שמאי על הנחייתם המסורה, להורי לאה ואברהם שחינכו וגידלו אותי באהבה ודאגה, למורים, חברים, מנהלים וקולגות רבים מהם למדתי שיעורים וכישורים רבי ערך ולאשתי הטרייה עירית על תמיכתה וחברתה.

אני מודה לטכניון על התמיכה הכספית הנדיבה בהשתלמותי.
המחקר נתמך על ידי הקרן הלאומית למדע (תקציב 1070/07) והאיחוד
האירופי במסגרת רשת FP7 למצויינות בתקשורת אלחוטית
(NEWCOM++).

תקציר

ראשיתה של תורת האינפורמציה בשנת 1948, עת פורסם מאמרו המכונן של C. E. Shannon [98]. במאמר זה הונחו היסודות התיאורתיים לשתי בעיות מהותיות בתורת האינפורמציה: בעיית קידוד המקור ובעיית קידוד הערוץ. התוצאה המרכזית הנלמדת מעבודתו המכוננת של Shannon היא כי בערוץ רועש ניתן להשיג תקשורת אמינה כרצוננו בקצבים הקרובים שרירותית לקיבול הערוץ.

תקשורת אמינה מושגת על ידי הוספת יתירות למידע המשודר בערוץ. תהליך זה נקרא קידוד ערוץ. היתירות המתווספת במהלך הקידוד, משמשת במקלט לצורך פענוח אמין של המידע המשודר בערוץ הרועש. Shannon ב- [98], השתמש בטכניקה המבוססת על קודי בלוק אקראיים. טכניקה זו הפכה למרכזית בתורת האינפורמציה, והיא משמשת במגוון רחב של מודלים ובעיות. יחד עם זאת, קוד בלוק אקראי איננו מעשי (על פי רוב, בעיקר משיקולים של סיבוכיות והשהיה). החיפוש אחר קודים המשיגים תקשורת אמינה בקצבים הקרובים לקיבול הערוץ והניתנים למימוש במערכות מעשיות מהווה כר נרחב למחקר בתורת הקודים מזה למעלה מ- 60 שנה.

מוקדי העניין העיקריים של ההחוקרים בתורת האינפורמציה כוללים את הנושאים הבאים:

- הבנת המגבלות היסודיות של סכמות קידוד במגוון של מודלים ובעיות.
- פיתוח והתאמת סכמות קידוד מבניות המשיגות את אותן מגבלות יסודיות.
- פיתוח ואנליזה של טכניקות אלגוריתמיות עבור סכמות קידוד מבניות.

בחלקו הראשון של המחקר המסוכם בעבודה זו, נלמדו כמה מהמגבלות היסודיות בתקשורת באמצעות קודים לא-בינאריים. בפרט, התמקדנו בגזירה של חסמים עליונים על הסתברות השגיאה בפענוח מסוג סבירות מירבית של קודי בלוק ליניאריים לא-בינאריים המשודרים בערוצים חסרי זיכרון וסימטריים. לאנליזה של הסתברות השגיאה בתקשורת מקודדת מוקדש חלק ניכר מהספרות המקצועית בשטח זה. הן המגבלות היסודיות על הסתברות השגיאה והן הערכת ביצועיהם של קודים מבניים ספציפיים נלמדו במגוון עבו-דות. כיוון שנדיר שביצועי מערכות תקשורת ניתנים להערכה על סמך ביטויים מדוייקים, נהוג להעריך את ביצועיהן של מערכות אלו באמצעות חסמים על הסתברות השגיאה. יתר

על כן, אנליזה של מערכות ספציפיות עלולה להיות בלתי מעשית. משום כך, מקובל לבחון צבירים של קודים ולהעריך את ביצועיהם על פי פרמטרים בסיסיים (כגון ספקטרום המרחקים הממוצע של הצביר). מערכות תקשורת מודרניות פועלות באמינות בקצבים הקרובים לקיבול בעוד שחסם האיחוד מתגלה כחסר ערך מעבר לקצב הקיטעון במערכות מקודדות בעלות אורך בלוק ממוצע עד גבוה. המגבלות של חסם האיחוד דירבנו את פיתוחן של טכניקות חסימה עדיפות. רובו של המחקר בשטח זה התמקד בקודי בלוק ליניאריים ובינאריים (ראה סקירה מקיפה ב-[94]).

הסימטריה המוגדרת בעבודתנו מרחיבה באופן טבעי את ההגדרה המקובלת בערוצים חסרי זיכרון, בינאריים-בכניסה וסימטריים-במוצא (MBIOS). עבור ערוצי MBIOS ידוע כי הסתברות השגיאה של קודי בלוק ליניאריים ובינאריים, תחת פענוח סבירות מירבית, איננה תלויה בהודעה המשודרת. תוצאה בסיסית זו שימשה לפיתוח מגוון רחב של חסמים עליונים על הסתברות השגיאה תחת פענוח סבירות מירבית של קודי בלוק לינאריים ובינאריים. בראשית מחקרנו הראנו כי אי-התלות בהודעה המשודרת מתקיימת גם עבור המקרה הלא-בינארי. החסמים העליונים שפותחו בעבודה זו שייכים למשפחה של חסמים בטכניקה של Gallager [45]. בפרט, המתקדנו בגרסה השנייה של החסמים של Duman ו- Salehi [26], [31], [32], [94] ו- [96]. החסמים שפותחו בחלק זה של המחקר משמשים להערכת הסתברות השגיאה של קודים בעלי מטריצת בדיקת זוגיות דלילה (LDPC). במחקרנו, בחנו את הסתברות השגיאה של קודי ה-LDPC הלא-בינאריים של Gallager [44] במגוון מודלים של ערוצי תקשורת (תחת פענוח סבירות מירבית). לשם כך, נתקבל הביטוי המדויק לספקטרום הקומפוזיציות המלא של קודים אלה (כהרחבה לפיתוח במקרה הבינארי ב-[18] ו- [105]). התוצאות שהתקבלו הושוו לחסמים תחתונים מסוג אריזת-כדורים (sphere-packing). מהשוואה זו ניתן ללמוד על היישימות של החסמים שפותחו בחלק זה של המחקר.

בחלקו השני של המחקר התמקדנו בביצועיהן של מערכות מקודדות עם מפענחים מוכללים. בכינוי מפענחים מוכללים הכוונה למפענחים הפועלים במצבים הבאים:

- המפענח רשאי לא לבצע החלטה על סמך האות הנקלט. הפלט של המפענח במצב זה מכונה מחיקה. כאשר החלטה בכל זאת נעשית ע"י המפענח, המאורע בו החלטת המפענח שגויה נקרא שגיאה בלתי מתגלית.
- המפענח רשאי להציע יותר מאשר החלטה בודדת בהסתמך על האות הנקלט. פלט המפענח במקרה זה נקרא רשימה (למפענח קוראים מפענח רשימה). המאורע בו הרשימה לא כוללת את ההודעה ששודרה נקרא שגיאת פענוח רשימה.

חוק ההחלטה האופטימאלי במצבים אלו פותח ע"י Forney יחד עם גזירתם של חסמים אקספוננציאלים על מאורעות השגיאה [41] (והגודל הממוצע של הרשימה, במקרים הרלוונטים). המפענח האופטימאלי שהוצע ע"י Forney נבדל מהותית ממפענח הרשימה עם אורך קבוע [36], [117]. החסמים האקספוננציאליים ב- [41] פותחו עבור קודי בלוק אקראיים.

חסמים שכאלו פותחו גם עבור מספר מפענחים תת-אופטימאליים (ראה למשל [9], [52] ו- [54]). המוטיבציה לעסוק במפענחים מוכללים נובעת בין היתר מהשימושים האפשריים של מפענחים אלו במערכות עם משוב ובמערכות עם קודים משורשרים. במערכות תקשורת הכוללות משוב, ניתן במקרים מסוימים להשיג שיפור ניכר בהסתברות השגיאה תוך שימוש במנגנונים לשידור מחדש של ההודעה.

במחקרנו עמדנו על ביצועיהם של המפענחים המוכללים הבאים:

- המפענח האופטימאלי של Forney, עבור המקרה של פענוח עם מחיקות ועבור המקרה של מפענח רשימה באורך משתנה.
- פענוח תת אופטימאלי מסוג likelihood ratio-test (LR) עבור פענוח עם מחיקות.
- פענוח עם רשימה באורך קבוע הכוללת את ההודעות הסבירות ביותר.

חסמים עליונים על הסתברויות מאורעות השגיאה תחת מפענחים מוכללים אלו פותחו עבור קודי בלוק ליניאריים (בינאריים ולא-בינאריים) המשודרים בערוצים חסרי זיכרון וסימטריים. חסמים אלו לוו בתוצאות על אי-תלות הסתברויות מאורעות השגיאה בהודעה המשודרת. החסמים מתאימים לקודים מבניים ספציפיים ולצבירים של ספרי קוד. תלות החסמים בקודים הנדונים באה לידי ביטוי דרך ספקטרום המרחקים (הממוצע) של הקוד (או צביר הקודים). הן ביצועים אסיפמטוטיים והן ביצועים עבור אורך בלוק סופי ניתנים לאנליזה באמצעות החסמים שפותחו. תוצאות המחקר מודגמות עבור קודי בלוק ליניאריים אקראיים (ביצועים אסימפטוטיים), ועבור צבירי קודים מסוג LDPC עם אורך בלוק סופי. שימושים לטובת האנליזה של ביצועיהן של מערכות עם סכמות לשידור חוזר אוטומטי (ARQ) מודגמים אף הם.

בחלקו השלישי של המחקר מוצעת מודיפיקציה של אלגוריתם Vitebri עבור מערכות מקודדות עם פענוח רשימה ופענוח עם משוב. אלגוריתם Vitebri וקודי קונבולוציה מהווים מרכיב מרכזי במערכות תקשורת מזה עשרות שנים. אלוגריתם Vitebri מאפשר לחשב פענוח סבירות מירבית תוך שמירה על סיבוכיות המתאימה למימוש במגוון בעיות בעיבוד אותות, בתורת הקודים ובתקשורת. קיימות מספר הכללות חשובות של אלגוריתם Vitebri, הן בתורת הקודים והן בתקשורת. בפרט, ראה הכללות של אלגוריתם Vitebri לפענוח רשימה ב- [20], [50], [70], [81], [88], [95] וב- [102], ועבור מערכות עם סכמות אוטומטיות לשידור חוזר ב- [51], [53], [67] וב- [120].

המודיפיקציה שפותחה בחלק זה של המחקר מתלכדת עם המפענח האופטימאלי של Forney עבור קודי קונבולוציה. עבודתנו מודגמת עבור פענוח עם מחיקות של קודי קונבולוציה. יחד עם זאת, המודיפיקציה מתאימה למקרה הכללי של תהליך מרקובי (Markov) עם מספר מצבים סופי הנצפה דרך ערוץ חסר זיכרון. ביצועיו של האלגוריתם שהצענו מושווים לביצועי פענוח תת-אופטימאלי מסוג LR, ולביצועיו של אלגוריתם תת-אופטימאלי נוסף אשר הוצע ע"י Yamamoto ו- Itoh ב- [120]. על אף שסכמת הפענוח ב- [120] פשוטה, ביצועיה קרובים מאוד לאלו של האלגוריתם האופטימאלי. ביצועיו של האלגוריתם מסוג LR נחותים מאלו המושגים בפענוח אופטימאלי.

בחלקו האחרון של המחקר התמקדנו באפליקציות של הטכניקה של קיטוב (polarization) של ערוצים. קודים קוטביים (polar codes) התגלו ע"י Arikan [4]. קודים אלו מהווים משפחה של קודים משיגי קיבול בערוצים בינאריים-בכניסה חסרי זיכרון וסימטריים-במוצא, תחת פענוח מסוג ביטול סידרתי (successive cancelation). התיאוריה והפרקטיקה של קודים קוטביים עדיין בתחילת דרכן. יחד עם זאת, כבר נמצאו מספר יישומים לטכניקה של קיטוב של ערוצים לבעיות קלאסיות בתקשורת ותורת האינפורמציה. יישומים למודלים בסיסיים במערכות מרובות משתמשים כגון ערוץ הפצה מדורג וערוץ מרובה גישה נלמדו ב- [62]. קודים קוטביים התגלו כאופטימאליים עבור דחיסת מקורות עם עיוות [62], [65]. יישומים של קודים קוטביים בבעיית Wyner-Ziv בינארית ובבעיית Gelfand-Pinsker בינארית נלמדו גם כן [65]. הקצבים האפשריים להשגה עם קודים קוטביים תחת פענוח ביטול סידרתי בבעיות עם ערוצי תרכובת (compound channels) נחקרו ב- [55]. מגוון אפשרויות אלו היווה את המוטיבציה לגילוי של יישומים אפשריים נוספים לטכניקה של קיטוב ערוצים. בחלק זה של המחקר הצענו סכמות המבוססות על קודים קוטביים לטובת תקשורת אמינה וחסויה בערוץ עם מאזין (wire-tap channel) ולטובת תקשורת בערוצים מקביליים עם סדר (permutation) מיפוי כניסה שרירותי.

ערוץ התקשורת עם מאזין הינו ערוץ מרובה משתמשים עם משדר בודד וזוג מקלטים. המקלט האחד משמש את המשתמש החוקי, והמקלט השני משמש את המאזין. נדרש בבעיה זו להשיג תקשורת אמינה וסודית עבור המשתמש הלגיטימי. החסם העליון ההדוק ביותר על הקצבים המאפשרים לקיים תקשורת זו הינו קיבול הערוץ עם סודיות. בחלק זה של המחקר הוכחנו כי תחת תנאים של ערוץ עם מאזין מדורג, בינארי בכניסה, חסר זיכרון וסימטרי במוצא, קיימים קודים מבניים מתאימים אשר משיגים את קיבול הערוץ עם סודיות.

בבעיה של ערוצים מקביליים עם סדר שרירותי בכניסה, התקשורת המקודדת מתנהלת בו זמנית דרך קבוצה נתונה של ערוצים במקביל. הקושי בבעיה זו נעוץ בכך שהמיפוי של מילות קוד לערוצים לא מוכתב ע"י המערכת אלא נבחר באופן שרירותי ע"י הערוץ. במחקרנו הנחנו כי סדר המיפוי נשמר קבוע לאורך התשדורת כולה וידוע למפענח במקלט. ניתן להראות כי קיבול הערוץ בבעיה זו שווה לסכום קיבולי הערוצים המקביליים (בתנאי שהמידה המשיגה את הקיבול בכל אחד מהערוצים הינה זהה בכל הערוצים). בחלק זה של המחקר הצענו סכמה המבוססת על קיטוב ערוצים עבור הבעיה הנדונה. הקוד הקוטבי שהצענו מבוסס בין היתר על קודים בעלי הפרדת מרחק מקסימאלית (MDS). עבור ערוצים מקביליים בינאריים בכניסה, חסרי זיכרון, סימטריים ומדורגים, הסכמה שהצענו משיגה את קיבול הערוץ. עבור המקרה בו הערוצים אינם מדורגים, הצענו חסמים תחתונים ועליונים על הקצבים האפשריים באמצעות הסכמה שהוצעה.