From Polar to Reed-Muller Codes: Unified Scaling, Non-standard Channels, and a Proven Conjecture

THÈSE N^{O} 7164 (2016)

PRÉSENTÉE LE 18 NOVEMBRE 2016

À LA FACULTÉ INFORMATIQUE ET COMMUNICATIONS
LABORATOIRE DE THÉORIE DES COMMUNICATIONS
PROGRAMME DOCTORAL EN INFORMATIQUE ET COMMUNICATIONS

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Marco MONDELLI

acceptée sur proposition du jury:

Prof. M. C. Gastpar, président du jury
Prof. R. Urbanke, directeur de thèse
Prof. A. Montanari, rapporteur
Prof. G. Kramer, rapporteur
Prof. H.-A. Loeliger, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE
Suisse
2016

A Mamma
A Papà

Sii avaro di citazioni. Diceva giustamente Emerson: "Odio le citazioni. Dimmi solo quello che sai tu".

Hold your quotes. Emerson aptly said: "I hate quotes. Tell me only what you know".

Umberto Eco, *La bustina di Minerva*

Acknowledgements

The man who said "I'd rather be lucky than good" saw deeply into life. One of the ways in which luck expresses itself is through the people we meet during our journey. I have been especially lucky in this regard, and this is the right time to acknowledge it.

I had the fortune, the honor and the absolute privilege to be advised by Rüdiger Urbanke. Much has been said about him and one way to write this paragraph would be to take a convex combination of the acknowledgements of his previous Ph.D. students. Such a combination would consist of the careful assembly of words like *generous*, *supportive*, *creative*, *persistent*, *fun-loving*, *wise*, *friend*, *guide*, *genius*, *hard worker*, *fast*, *optimistic*, and so on. This task is left as a simple stylistic exercise for the reader. Instead, I will focus a bit more on the last two adjectives of the list.

One lesson I learnt from Rüdiger is that 5 minutes is *some time*, 1 day is *a lot of time* and 3 days is *soooo much time!* For example, in 5 minutes you can eat lunch; in 1 day you can supervise an exam, grade it, scan it, generate a report summarizing the number of points obtained for every single problem, and send to each of the 350 students of the class his own report; and in 3 days... oh, in 3 days, you can solve a *very difficult problem*. Another lesson I learnt, and possibly the most important one, consists in the importance of always having a positive attitude in research and, in general, in life. To put it simply, if you think that you can solve a problem or achieve a goal, there is a chance that this will happen. Otherwise, it certainly will not.

I am grateful to the members of the thesis committee Michael Gastpar, Gerhard Kramer, Hans-Andrea Loeliger, and Andrea Montanari for their careful review of these pages and for their insightful comments. I would also like to thank Gerhard for hosting me for a few days at TUM in Munich, and Andrea for giving me the opportunity to collaborate with him in Stanford both in the past and in the forthcoming future.

During my Ph.D. studies I was fortunate to collaborate with great passionate researchers: Shrinivas Kudekar, Santhosh Kumar, Ivana Marić, Henry Pfister, Eren Şaşoğlu, and Igal Sason. I learnt a lot from each of them and I believe that these interactions helped me grow as a researcher.

The Information Processing Group (IPG) has been an amazing work environment and, ultimately, a kindergarten in which the kids can freely have (a lot of) fun, build things (typically bikes), and prove weird statements (but not too many) under benevolent adult supervision. I am sure that this place could have been the inspiration for countless *Peanuts* characters.

I need to thank a whole bunch of present and past members of IPG for making

my stay simply unforgettable. First, the more senior kids: Olivier Lévêque, Nicolas Macris, Bixio Rimoldi, and Emre Telatar. Olivier is the reasonable mathematician, abstract in thought and practical in life; Nicolas unveils the mysteries of the universe, but meanwhile gets lost; Bixio is the engineering sportsman and, when adult, wants to become a ski teacher; Emre just knows everything, but it is quite hard to find him.

Then, those who keep things running: Muriel Bardet, Françoise Behn, Giovanni Cangiani, and Damir Laurenzi. Before moving on, I would like to say a few more words about Muriel and Damir. Damir is the systems manager and, as a plus, cooks delicious cakes. He is meticulous, patient, and always happy to help at literally any time of the day/night. However, sometimes I have the impression that all these years spent fixing silly bugs (and the silliest, I admit, were mine) have added an intimidating note to his personality. For example, the following quote stands out on his office door: "Do not meddle in the affairs of sysadmins, for they are subtle and quick to anger". One day, he came to the office with the head completely shaved and I acknowledged his brand-new *Breaking Bad* haircut. That day all my processes running on the cluster mysteriously crashed. I sincerely hope that, when he reads these words, I will be sufficiently far from Switzerland. Muriel is the benevolent adult supervision. Suppose that you have a problem that is not about computers (for that, see Damir). Then, you should not try to fix it, but call Muriel, because she knows what to do. Among the uncountable occasions in which her help saved my day, I would like to mention two of them: the time when I was applying to rent my current place in Lausanne and the time when I thought I lost my bike+office+house keys during one of the several security checks at Tel Aviv airport.

Finally, the junior kids. I would like to thank Elie, Serj, and Wei for bearing me as an officemate; Marc B. (as in brilliant or, more simply, barefoot) for his loud and contagious good mood; Andrei, Stefano Rosati, and Young Jun for the biking, hiking, and via ferrata tours that *nearly* killed me (more specifically, I thank them for the *nearly*); Cyril for re-instilling in me the passion for running; Jean, Mohamad, and Rajai for the fun times at conferences and in Baghdad Café; Mani for the stark contrast between his extremely polite manners and his jokes; Rafah for accepting to hire me when I will be a hobo in the Bay Area. A heartfelt thank you goes to Hamed. He is an advisor (he guided me for most of my Ph.D.), a collaborator (five out of the seven technical chapters of this thesis are based on joint works with him), a former IPG member and, above all, a friend. He is brilliant, generous, selfless, and the list of adjectives could go on for much longer. One of the most amazing experiences of these Ph.D. years has been to go to his wedding. As an additional benefit, the wedding trip still provides me with a remarkable source of funny anecdotes to tell over dinner.

Thank you AJ, Marwa, Abbas, Renata, Sonia Bogos, Onur, Alexandros, Artem, and Christos for the Lausanne adventures. Thank you Jean-Louis, Stefano Olivotto, Alberto, Amos, and Roberto for all the lunches in the BC cafeteria (seriously? fish lasagna?) and for the memorable (and, at the same time, blurry) Christmas dinners. Thank you Meo, Mainak, Mahnoosh, Nariman, Sonia Bhaskar, Stefano Rini, and Yash for the great times in Stanford. Thank you Holly for reading my *whole* thesis, for adding corrections in 5 different colors (is a green error better than a blue one?) and, ultimately, for listening to my objections to those corrections without kicking me out of your office.

The Ph.D. life (and, I guess, life in general) is not really a straight path, but more of a roller coaster ride. Hence, Ladies and Gentlemen, I would like to introduce the prestigious *Save My Rear End Award* whose purpose is to recognize up to two outstanding individuals that helped me in a moment of exceptional difficulty. The Award consists of a certificate and a rich honorarium of $1. And the winners are... *Meo* for her fundamental contributions in psychology exemplified by the sentence "Marco, you should get out of here. This lady is nuts.", which saved me from a crazy landlord; and *Francesco* for mastering the practice of hospitality when I showed up in his office in Liège, basically unannounced and in a state of particular distress. Congratulations Meo and Francesco! Enjoy your dollar, as you truly deserve it!

A special thank you goes to Chiara because, even if things end, memories live forever.

Despite my best efforts, English remains a foreign language and there are some concepts that simply do not feel right when translated. For this reason, the final part of these acknowledgements is in *Italian*. Un enorme grazie va a un piccolo gruppo di persone che ho conosciuto durante il liceo e l'università. In particolare, vorrei ringraziare Alberto, perchè ha sempre la stessa soluzione ad ogni problema; Alessandro, per le mille risate sparse nei passati tre lustri; Annamaria, perchè "di perle come te ce ne sono poche"; Fabio, perchè sono passati altri quattro anni, ma è sempre come se fossimo vicini di banco al liceo; Francesco, ed il perchè è contenuto qualche riga sopra; Giulia, perchè Atlanta è piuttosto tremenda ma quell'esperienza è stata indimenticabile; Marco, altrimenti noto come "Schicchi", per le mitiche serate pisane; Tommaso, perchè credo sia la persona che riesca a capirmi meglio.

Questa tesi è dedicata ai miei genitori. Dire che i miei ringraziamenti più sentiti vanno a loro sembra quasi superfluo. Ci sono stati sempre, ci saranno sempre. Da nove anni non vivo più a Siena, ma lì è rimasta *casa*.

Lausanne, 3 November 2016 M. M.

Abstract

The year 2016, in which I am writing these words, marks the centenary of Claude Shannon, the father of information theory. In his landmark 1948 paper "A Mathematical Theory of Communication", Shannon established the largest rate at which reliable communication is possible, and he referred to it as the *channel capacity*. Since then, researchers have focused on the design of practical coding schemes that could approach such a limit. The road to channel capacity has been almost 70 years long and, after many ideas, occasional detours, and some rediscoveries, it has culminated in the description of low-complexity and provably capacity-achieving coding schemes, namely, polar codes and iterative codes based on sparse graphs. However, next-generation communication systems require an unprecedented performance improvement and the number of transmission settings relevant in applications is rapidly increasing. Hence, although Shannon's limit seems finally close at hand, new challenges are just around the corner.

In this thesis, we trace a road that goes from polar to Reed-Muller codes and, by doing so, we investigate three main topics: unified scaling, non-standard channels, and capacity via symmetry.

First, we consider *unified scaling*. A coding scheme is capacity-achieving when, for any rate smaller than capacity, the error probability tends to 0 as the block length becomes increasingly larger. However, the practitioner is often interested in more specific questions such as, "How much do we need to increase the block length in order to halve the gap between rate and capacity?". We focus our analysis on polar codes and develop a unified framework to rigorously analyze the scaling of the main parameters, i.e., block length, rate, error probability, and channel quality. Furthermore, in light of the recent success of a list decoding algorithm for polar codes, we provide scaling results on the performance of list decoders.

Next, we deal with *non-standard channels*. When we say that a coding scheme achieves capacity, we typically consider binary memoryless symmetric channels. However, practical transmission scenarios often involve more complicated settings. For example, the downlink of a cellular system is modeled as a broadcast channel, and the communication on fiber links is inherently asymmetric. We propose provably optimal low-complexity solutions for these settings. In particular, we present a polar coding scheme that achieves the best known rate region for the broadcast channel, and we describe three paradigms to achieve the capacity of asymmetric channels. To do so, we develop general coding "primitives", such as the chaining construction that has already proved to be useful in a variety of communication problems.

Finally, we show how to achieve *capacity via symmetry*. In the early days of

coding theory, a popular paradigm consisted in exploiting the structure of algebraic codes to devise practical decoding algorithms. However, proving the optimality of such coding schemes remained an elusive goal. In particular, the conjecture that Reed-Muller codes achieve capacity dates back to the 1960s. We solve this open problem by showing that Reed-Muller codes and, in general, codes with sufficient symmetry are capacity-achieving over erasure channels under optimal MAP decoding. As the proof does not rely on the precise structure of the codes, we are able to show that symmetry alone guarantees optimal performance.

Keywords: asymmetric channel, broadcast channel, capacity-achieving codes, capacity via symmetry, chaining construction, list decoding, polar codes, Reed-Muller codes, scaling, sparse graph codes.

Abstract

L'anno 2016, in cui scrivo queste parole, segna il centenario della nascita di Claude Shannon, il padre della teoria dell'informazione. Nel suo fondamentale articolo del 1948 "A Mathematical Theory of Communication", Shannon stabilì il massimo tasso di informazione che può essere trasmesso garantendo una comunicazione affidabile e lo definì *capacità di canale*. Da allora, i ricercatori si sono concentrati sulla progettazione di sistemi pratici di codifica che potessero avvicinarsi ad un tale limite. Il cammino verso la capacità è durato quasi 70 anni e, dopo numerose idee, deviazioni occasionali e qualche riscoperta, è culminato nella descrizione di schemi di codifica a bassa complessità e capaci di raggiungere la capacità, ossia i codici polari e i codici iterativi basati su grafi sparsi. Tuttavia, i sistemi di comunicazione di nuova generazione richiedono un miglioramento delle prestazioni senza precedenti ed il numero di scenari trasmissivi rilevanti nelle applicazioni sta crescendo rapidamente. Quindi, anche se il limite di Shannon pare ormai raggiunto, nuove sfide sono dietro l'angolo.

In questa tesi, tracciamo un cammino che va dai codici polari a quelli di Reed-Muller e, così facendo, investighiamo tre argomenti principali: leggi di scala unificate, canali non-standard e capacità attraverso la simmetria.

Innanzi tutto, consideriamo le *leggi di scala unificate*. Un sistema di codifica raggiunge la capacità quando, per ogni tasso di informazione trasmesso inferiore alla capacità, la probabilità di errore tende a 0 all'aumentare della lunghezza di blocco. Tuttavia, da un punto di vista pratico, si è spesso interessati a rispondere a domande specifiche quali, "Di quanto dobbiamo aumentare la lunghezza di blocco in modo da dimezzare la differenza tra tasso di informazione trasmesso e capacità?". In questa tesi, ci concentriamo sui codici polari e sviluppiamo una teoria unificata per analizzare in modo rigoroso le leggi di scala dei parametri fondamentali, ossia la lunghezza di blocco, il tasso di informazione trasmesso, la probabilità di errore e la qualità del canale. Inoltre, alla luce del recente successo di un algoritmo di decodifica con lista per i codici polari, presentiamo dei risultati sulle leggi di scala che regolano le prestazioni dei decodificatori con lista.

Successivamente, discutiamo i *canali non-standard*. Quando diciamo che un sistema di codifica raggiunge la capacità, consideriamo tipicamente canali binari, privi di memoria e simmetrici. Tuttavia, in pratica si osservano spesso degli scenari più complicati. Ad esempio, il downlink di un collegamento cellulare viene modellato come un canale broadcast e la comunicazione su fibra ottica è intrinsecamente asimmetrica. In questa tesi, proponiamo e dimostriamo l'ottimalità di una serie di soluzioni a bassa complessità adatte a questi scenari. In particolare, presentiamo un sistema di codifica polare che permette di raggiungere i tassi di informazione trasmessi ottimi noti per il canale broadcast e descriviamo tre paradigmi di codi-

fica capaci di raggiungere la capacità dei canali asimmetrici. Per ottenere questo obiettivo, sviluppiamo delle generiche "primitive" di codifica, come una costruzione a catena che si è già dimostrata utile in una varietà di problemi di comunicazione.

Infine, dimostriamo come raggiungere la *capacità attraverso la simmetria*. Nei primi anni della teoria dei codici, una tecnica popolare consisteva nello sfruttare la struttura dei codici algebrici per ideare algoritmi pratici di decodifica. Tuttavia, dimostrare l'ottimalità di tali schemi di codifica rimase un obiettivo sfuggente. In particolare, la congettura che i codici di Reed-Muller raggiungono la capacità risale agli anni '60. In questa tesi, risolviamo questo problema dimostrando che i codici di Reed-Muller e, in generale, codici dotati di sufficiente simmetria raggiungono la capacità su canali a *erasure* con decodifica ottima MAP. Dal momento che la dimostrazione non dipende dalla struttura precisa dei codici, siamo in grado di provare che la simmetria da sola garantisce prestazioni ottime.

Parole chiave: canale asimmetrico, canale broadcast, codici che raggiungono la capacità, capacità attraverso la simmetria, costruzione a catena, decodificatore con lista, codici polari, codici di Reed-Muller, leggi di scala, codici basati su grafi sparsi.

Contents

List of Figures

1 Introduction

Nomina direttamente autori e personaggi di cui parli, senza perifrasi. Così faceva il maggior scrittore lombardo del XIX secolo, l'autore del 5 maggio.

Name the authors and characters you refer to, without using periphrases. So did the greatest Lombard author of the 19th century, the author of "The 5th of May".

On 18 October 1989, the *Galileo* spacecraft began its journey towards Jupiter to send off a probe that would accomplish the most difficult atmospheric entry ever attempted. However, on 11 April 1991, when the antenna for data transmission had to be deployed, no confirmation signal came back. Despite being a huge setback, this event did not cause the failure of the mission: the Jet Propulsion Lab reprogrammed into the spacecraft computers an elaborate coding scheme able to operate with a 2 dB gap to channel capacity and with an error probability of the order of 10^{-7}, which arguably represented the highest-performance and highest-complexity system for error control existing at the time (see Chapter 3.4 of [1] for more details).

This is possibly the most spectacular case in which coding theory saved the day. On a more mundane level, each of us uses the fruits of coding theory every day when we pick up our cell phones, transfer information over the Internet, or store files on our computers. As a preliminary example, just consider the two classes of codes that constitute the main focus of this thesis: a first-order *Reed-Muller code* was employed in the *Mariner* Mars 1969 and 1971 spacecraft, and *polar codes* are among the most promising candidates for the incoming 5G standard.

To establish the context, we begin with a quick history of coding theory in Section 1.1. Then, we move to the present and, in Section 1.2, we briefly describe the main research topics investigated in this work. After this general discussion, we present in some more detail the various components that constitute the title of this thesis: in Section 1.3, we deal with polar codes; in Section 1.4, we discuss Reed-Muller codes

and a rather old conjecture (spoiler alert); in Section 1.5, we talk about scaling; in Section 1.6, we consider coding for non-standard channels. Finally, in Section 1.7, we briefly summarize the contributions of this thesis.

Right after the title and at the beginning of each chapter, there is a quote containing a piece of advice on how to write well. Indeed, I spent quite some time thinking about the appropriate style for this thesis and, after discovering a fairly long list of recommendations from a witty Italian semiologist, I decided to reproduce some of them. I have to admit that I did not follow thoroughly all these rules, but I hope that the reader can find some value in them.

1.1 Channel Coding: The Shortcut to Channel Capacity

In this thesis, we study the problem of communicating through a noisy channel W. For the moment, let us consider the simplest instance of such a problem: there is a single transmitter who wishes to communicate with a single receiver, and the noisy channel is binary, memoryless and symmetric (BMS). In order to be even more concrete, take the case of the binary erasure channel (BEC). This channel erases each of the inputs independently with probability ε and, otherwise, leaves the input as is. When the binary input $x \in \{0, 1\}$ is erased, the corresponding output y is denoted by a question mark "?". The BEC is perhaps the simplest non-trivial channel model that can be imagined, and it was introduced by Elias as a toy example in 1954 [2]. In more recent times, the erasure channel has been promoted to the class of "real-world" channels, as it can be used to model data networks, where packets either arrive correctly or are lost due to buffer overflows or excessive delays. Our interest, however, in the erasure channel is mainly due to the combination of the following two facts: its simplicity makes the theoretical analysis significantly easier; and, quite surprisingly, many of the properties that hold for the BEC turn out to be true in much greater generality. In a nutshell, the idea is to prove a result first for the BEC and, later on, attack more general scenarios. This philosophy proved to be extremely successful for the case of low-density parity-check (LDPC) codes and of polar codes, i.e., for two of the most popular classes of codes today.

More specifically, the transmitter wants to send to the receiver K bits of information, i.e., a sequence $u_{1:K} = (u_1, \dots, u_K) \in \{0, 1\}^K$. The problem is that the channel W is noisy; for example, some of the inputs might be erased. Hence, in order to improve the reliability of the communication, the transmitter adds some redundancy and sends a total of N binary symbols, i.e., a sequence $x_{1:N} = (x_1, \dots, x_N) \in \{0, 1\}^N$. In other words, the information sequence $u_{1:K}$ is mapped into a codeword $x_{1:N}$ via the encoder $\mathcal{E}$. The set of codewords that are associated with all possible information sequences represents the code $\mathcal{C}$. The block length of the code is N and the rate R is given by the ratio between the number of information bits and the block length, i.e., $R = K/N$. On the receiver side, the sequence $y_{1:N} = (y_1, \dots, y_N) \in \{0, 1\}^N$ obtained as output from the channel is processed by the decoder $\mathcal{D}$, in order to obtain an estimate $\hat{u}_{1:K} = (\hat{u}_1, \dots, \hat{u}_K) \in \{0, 1\}^K$ of the original information sequence. The block error probability P_{B} is the probability that the information sequence is reconstructed correctly. The situation is schematically represented in Figure 1.1.

In his landmark 1948 paper [3], Shannon defines a fundamental quantity called

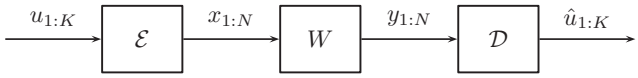


Figure 1.1 – Basic communication scenario. The information sequence $u_{1:K}$ is mapped into the codeword $x_{1:N}$ by the encoder $\mathcal{E}$. The channel W outputs the noisy sequence $y_{1:N}$ that is used by the decoder $\mathcal{D}$ to obtain the estimate $\hat{u}_{1:K}$ of the actual information sequence.

the *channel capacity*. This quantity represents the maximum possible rate at which reliable transmission is possible, when N becomes larger and larger. In formulae, for any $R < C(W)$, where $C(W)$ denotes the capacity of the channel W, there exists a family of codes $\{\mathcal{C}_N\}$ with block length N and rates R_N converging to R such that

$$\lim_{N\to\infty} P_{\mathrm{B}}(N, R, W) = 0. \tag{1.1}$$

Conversely, for any $R > C(W)$, the error probability tends to 1 for any family of codes.

Shannon's achievability argument was based on *random coding*. The central idea is to create an ensemble of codes and then to study the error probability of a typicality decoder. Basically, if the ensemble has pairwise independent codewords, then the resulting family of codes is capacity-achieving. Classic proofs of Shannon's channel coding theorem can be found in [4–7]. A related technique consists in looking directly at the *weight distribution* of the codes, i.e., at the distribution of the number of 1s of the codewords. If this weight distribution is sufficiently close to the weight distribution of a random ensemble, then it is possible to show that the family of codes is capacity-achieving [8, 9]. Unfortunately, random codes could not be used in practical applications, as their decoding complexity is exponential in the block length. As a result, the quest for practical codes that could achieve channel capacity had officially begun!

An excellent review on the history of coding theory can be found in [10]. As the most careful readers might have noticed, the survey [10] also inspired the title of the current section that does not, however, have such an ambitious goal. Indeed, we will focus only on the three coding paradigms that are connected to the body of this thesis: algebraic coding, iterative coding (or codes based on sparse graphs) and polar coding.

The *algebraic coding* paradigm dominated the first decades of coding theory [11, 12]. The central idea is to take advantage of the algebraic structure and the symmetries of the codes, in order to devise efficient encoding and decoding algorithms. This approach was quite different from Shannon's original one. Indeed, Shannon's argument is probabilistic and it shows that a certain family of codes is good in the typical case. On the contrary, algebraic codes have a deterministic construction. For this reason, they are more suitable for a worst-case type of analysis. Instead of proving that an ensemble of codes is good on average, the purpose is to show that a code guarantees a reliable transmission up to an assigned noise level in the channel (e.g., a fixed number of errors or erasures). Hence, the principal objective of algebraic coding theory is to maximize the separation between codewords, i.e., to design

codes with a minimum distance as large as possible. Examples of popular algebraic codes include Hamming [13], Golay [14], BCH [15,16] and Reed-Solomon [17] codes. Also the Reed-Muller codes [18,19] mentioned earlier belong to this class of codes.

Iterative coding was born in 1993 at the IEEE International Conference on Communications in Geneva, Switzerland, where Berrou, Glavieux, and Thitimajshima presented a new class of "turbo" codes capable of achieving near-Shannon-limit performance with reasonable decoding complexity [20]. Shortly thereafter, Gallager's LDPC codes [21] were rediscovered independently by MacKay [22] and Spielman [23], along with a low-complexity iterative decoder. Wiberg, Loeliger, and Kötter showed in [24,25] that the iterative decoding algorithms of both turbo and LDPC codes are instances of a general belief-propagation (BP) algorithm, and that turbo and LDPC codes themselves fall under the umbrella of *codes based on sparse graphs*. In this way, they rediscovered several results originally described in Tanner's largely forgotten paper [26]. A generic message-passing algorithm, namely, the sum-product algorithm, was presented by Kschischang, Frey, and Loeliger [27], in order to compute marginal functions in a Tanner graph or, as it is more commonly called, a "factor graph". For a tutorial on factor graphs, the interested reader is referred to [28]. In 2001, Luby *et al.* showed that sequences of irregular LDPC codes are capacity-achieving on the BEC [29]. Furthermore, a new powerful technique, called density evolution, was developed by Richardson and Urbanke, in order to prove that error-free performance could be achieved below a certain noise threshold for long codes and large numbers of iterations [30,31]. Using this approach, it was possible to optimize the degree distribution of irregular codes. In particular, Chung *et al.* designed several rate-1/2 codes for the additive white Gaussian noise (AWGN) channel, including one whose theoretical threshold approached the Shannon limit within 0.0045 dB [32]. Convolutional LDPC codes, also known as spatially coupled LDPC codes, were introduced by Felström and Zigangirov in [33]. Lentmaier *et al.* introduced a terminated version of convolutional LDPC ensembles, considered their density evolution analysis, and determined the thresholds for the BEC and for the binary-input AWGN channel [34]. Eventually, Kudekar, Richardson, and Urbanke proved in [35,36] that spatially coupled LDPC codes, decoded with low-complexity belief-propagation algorithms, achieve capacity universally over the class of BMS channels.

A completely different approach to the problem of achieving channel capacity is provided by *polar coding*. In his seminal 2009 paper [37], Arıkan discovered the technique of channel polarization and proved that polar codes are capacity-achieving for any BMS channel with low encoding and decoding complexity. It was recently pointed out in [38] that these codes were already considered by Stolte in his 2002 Ph.D. thesis [39], where he focused on codes generated by the Plotkin construction. However, Stolte did not conjecture that such codes were capacity-achieving. Contrarily to codes based on sparse graphs, polar codes have a deterministic construction. Contrarily to algebraic codes, this deterministic construction does not come from any explicit symmetry or structure in the code, rather from the process of channel polarization. Furthermore, the decoding algorithm for polar codes operates in a successive fashion with a single pass on the data, as opposed to iterative decoding. Since their introduction, polar codes have become a very popular subject both in academia and in the industry, as testified by the ever growing number of papers and patents based on them.

In summary, the main techniques to achieve channel capacity are as follows:

Random coding and weight distribution. If the codewords have a uniform distribution and are pairwise independent, then the family of codes achieves capacity. In fact, we only need that the weight distribution of the family of codes is sufficiently close to the weight distribution of random codes. On the downside, no efficient decoding algorithm (anything with complexity smaller than exponential in the block length) is known for the transmission of these codes over general channels.

Iterative coding on sparse graphs. Spatially coupled LDPC codes achieve the capacity of any binary memoryless symmetric channel and can be decoded with a low-complexity belief-propagation algorithm. The idea of the proof consists in studying the density evolution of the decoding process, in order to establish the threshold at which reliable communication is possible.

Polar coding. Polar codes achieve the capacity of any binary memoryless symmetric channel and can be decoded with a low-complexity successive cancellation algorithm. The proof is somehow "baked" into the construction of the codes and it is based on the general phenomenon of channel polarization. The idea of polarization is to transform identical copies of the transmission channel into either completely noisy channels or completely noiseless channels, while conserving the overall capacity.

1.2 What Now?

Paraphrasing an infamous talk at the first IEEE Communication Theory Workshop in 1971, the survey [10] ends by asking whether coding theory is finally dead. Unsurprisingly, the conclusion is that coding theory is alive and quite well. Perhaps, the best proof of this fact comes from the following observation: the only two classes of codes that are provably capacity-achieving for any BMS channel with affordable complexity, i.e., polar codes and spatially coupled codes, were introduced *after* the publication of the aforementioned survey.

Despite that Shannon's limit seems finally close at hand, new questions and challenges are just around the corner, due to the unprecedented performance improvement required by next-generation systems and the increasing number of communication scenarios relevant in practical applications. In this thesis, we provide several original contributions concerning three main topics, namely, unified scaling, non-standard channels, and capacity via symmetry.

Unified scaling. When we consider the transmission over a noisy channel by using a coding scheme, the parameters of interest are as follows: the rate R that represents the amount of information transmitted per channel use, the block length N that represents the total number of channel uses, the block error probability P_{B}, and the quality of the channel W that can be quantified, e.g., by its capacity $C(W)$. We say that a family of codes achieves capacity when, for any $R < C(W)$, $P_{\mathrm{B}} \to 0$ as $N \to \infty$, see also (1.1). This means that, as we use increasingly longer block lengths, we can communicate with a vanishing error probability at the highest possible rate. However, when designing a

practical communication system, in principle we would be interested in characterizing exactly the relationship of R, N, P_{B}, and the quality of the channel W. Clearly, this is a formidable task. Hence, the problem has been tackled by studying the *scaling* of these parameters in various regimes. This would enable the practitioner to answer questions such as, "How much does the block length N need to increase, in order to halve the gap to capacity $C(W) - R$?", and "How much does the quality of the channel W need to improve, in order to make the error probability P_{B} decrease by a factor of 10?". In Chapter 2, we present several new results for the scaling of polar codes. In particular, we are able to provide a *unified characterization* of the relation between the relevant parameters in the *finite-length analysis of polar codes* under successive cancellation decoding. The next natural question is how to improve the scaling of polar codes. One promising candidate in this regard seems to be list decoding that yields excellent results in numerical simulations. In Chapter 3, however, we prove some negative results. We provide bounds on the performance of list decoders, and we show that a *list of finite size does not* suffice to *improve the speed at which capacity is approached*.

Non-standard channels. In the middle of the iterative coding revolution, during his 2001 plenary talk at the IEEE International Symposium on Information Theory, McEliece asked: "Are turbo-like codes effective on nonstandard channels?" [40]. At the time, the excellent performance of such codes on binary memoryless symmetric channels was already well established. Hence, it made sense to investigate how to use them in more general scenarios, i.e., for the transmission over channels that are non-symmetric, non-binary, non-memoryless, and multi-user. The basic conclusion of McEliece was that binary turbo-like codes with graph-based iterative decoding show a great potential on basically all such channels. In the last few years, coding theory has seen extraordinary advances that have led to the discovery of polar codes and spatially coupled LDPC codes. These two classes of codes are capable of provably achieving the capacity of the whole class of BMS channels with low-complexity algorithms. As a result, today more than ever, it is of fundamental importance to investigate to what degree the results on standard channels can be extended to non-standard ones. Indeed, it is no mystery that the next-generation communication systems call for order-of-magnitude increases in connectivity, speed, and data volume. In order to achieve such an ambitious goal, we need more efficient schemes that can deal not only with simple point-to-point symmetric channels, but also with non-standard scenarios. In Chapter 4, we show a novel *low-complexity construction based on polar codes* that *provably achieves the best known rate region for the broadcast channel*, i.e., Marton's region. In Chapter 5, we present and compare *three coding paradigms for achieving efficiently the capacity of asymmetric channels*. We regard these paradigms as general "meta-schemes", that provide many specific coding solutions, some old and some new, suitable to a variety of asymmetric settings.

Capacity via symmetry. As described at the end of the previous section, the main techniques for achieving channel capacity involve randomness, polariza-

tion, and iterative codes on sparse graphs. Whereas, a very popular paradigm in the first decades of the history of coding theory consisted in exploiting the algebraic structure of the codes, in order to design low-complexity encoding and decoding schemes. However, proving that such classes of codes were asymptotically optimal seemed, if not impossible, a rather elusive goal. More specifically, the conjecture that Reed-Muller codes are capacity-achieving was first discussed at the end of the 1960s. In more recent times, the interest in the problem was revived by the similarities between Reed-Muller codes and polar codes. In Chapter 7, we *solve this long-standing open problem*: we prove that Reed-Muller codes and, in general, *codes with sufficient amount of symmetry are capacity-achieving* under MAP decoding for the transmission over the BEC. In Chapter 8, we discuss some *generalizations* of this result. In particular, by carefully bounding the weight distribution, we show how to tighten results on the bit error probability to the block error probability.

1.3 Channel Polarization and Polar Codes

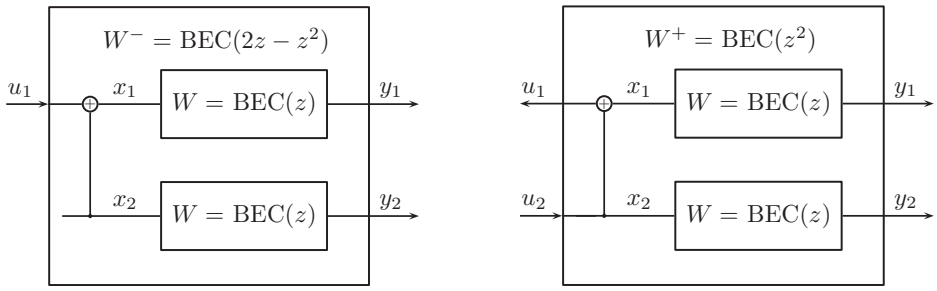
The purpose of this section is to get an idea of polar codes and of channel polarization. In this way, we introduce the notation and the concepts that will be useful throughout the thesis. As several excellent reviews already exist (see, for example, [41], Chapter 2 of [42] and Chapter 2 of [43]), we will be quite telegraphic.

1.3.1 Polarization Process

There are two types of binary channels for which it is easy to construct capacity-achieving codes. On the one hand, consider a completely noisy channel that sends any input to the same output symbol. This channel has a capacity of 0 bits. As no reliable communication is possible, sending uncoded bits is optimal. On the other hand, consider a completely noiseless channel, in which the output is always equal to the input. This channel has a capacity of 1 bit. As there is no noise, sending uncoded bits is optimal also in this case. The idea of channel polarization consists in taking independent copies of the transmission channel and transforming them into a set of completely noiseless channels and a set of completely noisy channels, in such a way that the overall capacity is preserved.

In order to formalize this idea, consider first the case of the erasure channel. Let W be a BEC(z), where z is the erasure probability. The capacity of this channel is $C(W) = 1 - z$. Suppose that you want to transmit two information bits and denote them by u_1 and u_2. Take two independent copies of W and send as inputs $x_1 = u_1 \oplus u_2$ and $x_2 = u_2$, where $\oplus$ denotes the XOR operator. Once the channel outputs y_1 and y_2 have been received, decode the information bits in a successive fashion: first, decode u_1 assuming no information about u_2 (i.e., by treating u_2 as noise); then, decode u_2 assuming that the value of u_1 is known.

As u_2 is treated as noise when estimating u_1, it is possible to recover u_1 if and only if both the inputs x_1 and x_2 are not erased. Now, assume that the value of u_1 is known. Then, it is possible to recover u_2 if and only if at least one of the inputs x_1 and x_2 is not erased. As a result, the channel W^- "seen" by u_1 is an erasure channel with erasure probability $2z - z^2$, and, given u_1, the channel W^+ "seen" by u_2 is an erasure channel with erasure probability z^2. Note that W^- is a worse



(a) Synthetic channel W^- "seen" by u_1. (b) Synthetic channel W^+ "seen" by u_2.

Figure 1.2 – One step of channel polarization: two copies of the original channel $W = \text{BEC}(z)$ are transformed into a worse channel $W^- = \text{BEC}(2z - z^2)$ and a better channel $W^+ = \text{BEC}(z^2)$ such that $C(W^-) + C(W^+) = 2C(W)$.

channel than W, and W^+ is a better channel than W, because $z^2 \leq z \leq 2z - z^2$. In addition, the sum of the capacities of the two synthetic channels W^- and W^+ is equal to the sum of the capacities of the two copies of the original channel W, i.e., $C(W^-) + C(W^+) = 2C(W)$. In other words, we have transformed two i.i.d. copies of the original transmission channel $W = \text{BEC}(z)$ into a worse channel $W^- = \text{BEC}(2z - z^2)$ and a better channel $W^+ = \text{BEC}(z^2)$, such that the sum of the capacities stays preserved under this transformation (see also Figure 1.2).

The procedure described above constitutes one step of polarization. In the second step of polarization, we transform two i.i.d. copies of W^- into a worse channel $W^{(-,-)}$, which is a BEC with erasure probability $2(2z-z^2)-(2z-z^2)^2 = 1-(1-z)^4$, and a better channel $W^{(-,+)}$, which is a BEC with erasure probability $(2z - z^2)^2$. Similarly, from W^+ we obtain the worse channel $W^{(+,-)}$, which is a BEC with erasure probability $2z^2 - z^4$, and the better channel $W^{(+,+)}$, which is a BEC with erasure probability z^4. It is easy to check that

$$C(W^{(-,-)}) + C(W^{(-,+)}) + C(W^{(+,-)}) + C(W^{(+,+)}) = 4C(W).$$

After n steps of polarization, we obtain 2^n synthetic channels that can be indexed as $W_n^{(i)}$, for $i \in \{1, \cdots, 2^n\}$. The channels $\{W_n^{(i)}\}$ are all erasure channels whose erasure probabilities are obtained as follows. Consider the random process Z_n, defined recursively as

$$Z_n = \begin{cases} 2Z_{n-1} - Z_{n-1}^2, & \text{w.p. } 1/2, \\ Z_{n-1}^2, & \text{w.p. } 1/2, \end{cases} \tag{1.2}$$

with $Z_0 = z$. Then, Z_n assumes with uniform probability 2^n distinct values that represent the erasure probabilities of the 2^n channels $\{W_n^{(i)}\}$. The situation is schematized in Figure 1.3.

Now, the synthetic channels $\{W_n^{(i)}\}$ polarize in the sense that Z_n converges almost surely to $Z_\infty \in \{0, 1\}$, as $n \to \infty$. Furthermore,

$$\sum_{i=1}^{2^n} C(W_n^{(i)}) = 2^n C(W).$$

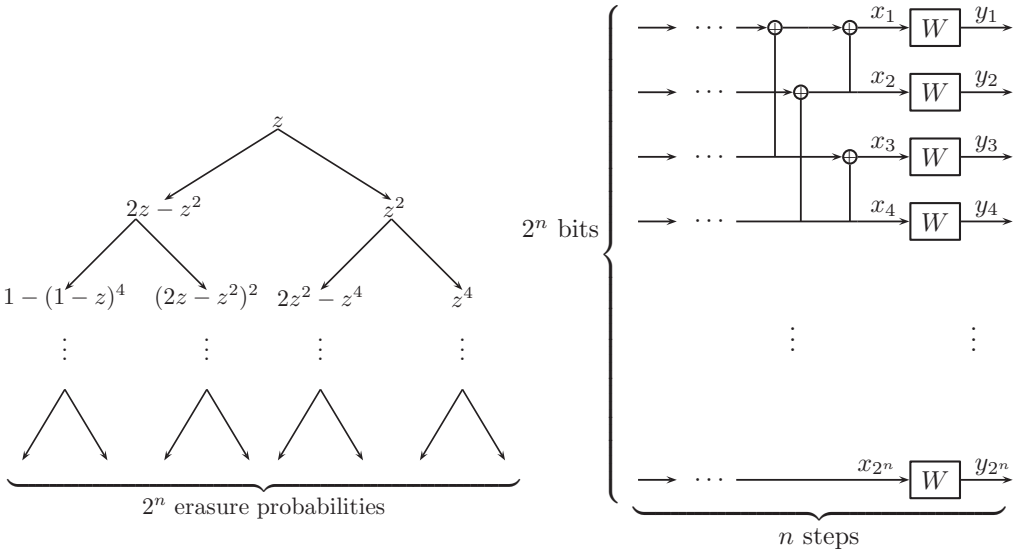


Figure 1.3 – n steps of channel polarization: 2^n copies of the original channel $W = \text{BEC}(z)$ are transformed into the 2^n synthetic channels $\{W_n^{(i)}\}$ that are all erasure channels such that $\sum_{i=1}^{2^n} C(W_n^{(i)}) = 2^n C(W)$.

Hence, a fraction $C(W)$ of the polarized channels is noiseless, i.e., it has capacity 1, and the remaining fraction $1 - C(W)$ is completely noisy, i.e., it has capacity 0.

In general, let W be a BMS channel with input alphabet $\mathcal{X} = \{0, 1\}$ and output alphabet $\mathcal{Y}$, and let $\{W(y \mid x) : x \in \mathcal{X}, y \in \mathcal{Y}\}$ be the transition probabilities. Denote by $C(W) \in [0, 1]$ the capacity of W. In order to quantify the reliability of channel, we use the Bhattacharyya parameter of W, denoted by $Z(W) \in [0, 1]$, that is defined as

$$Z(W) = \sum_{y \in \mathcal{Y}} \sqrt{W(y \mid 0) W(y \mid 1)}. \tag{1.3}$$

Note that if W is a BEC(z), then its Bhattacharyya parameter equals the erasure probability, i.e., $Z(W) = z$. The Bhattacharyya parameter $Z(W)$ is related to the capacity $C(W)$ via

$$Z(W) + C(W) \geq 1, \tag{1.4}$$
$$Z(W)^2 + C(W)^2 \leq 1, \tag{1.5}$$

both proven in [37].

The basis of channel polarization consists in mapping two identical copies of the channel $W : \mathcal{X} \to \mathcal{Y}$ into the pair of channels $W^- : \mathcal{X} \to \mathcal{Y}^2$ and $W^+ : \mathcal{X} \to \mathcal{X} \times \mathcal{Y}^2$, defined as

$$\begin{aligned} W^-(y_1, y_2 \mid x_1) &= \sum_{x_2 \in \mathcal{X}} \frac{1}{2} W(y_1 \mid x_1 \oplus x_2) W(y_2 \mid x_2), \\ W^+(y_1, y_2, x_1 \mid x_2) &= \frac{1}{2} W(y_1 \mid x_1 \oplus x_2) W(y_2 \mid x_2). \end{aligned} \tag{1.6}$$

Then, W^- is a worse channel, and W^+ is a better channel than W. This statement can be quantified by computing the relations among the Bhattacharyya parameters of W, W^- and W^+:

$$Z(W)\sqrt{2 - Z(W)^2} \leq Z(W^-) \leq 2Z(W) - Z(W)^2, \tag{1.7}$$

$$Z(W^+) = Z(W)^2, \tag{1.8}$$

which follow from Proposition 5 of [37] and from Exercise 4.62 of [44]. As previously pointed out, if W is a BEC, then also W^- and W^+ are BECs such that (1.8) still holds and $Z(W^-)$ is as large as possible, i.e.,

$$Z(W^-) = 2Z(W) - Z(W)^2. \tag{1.9}$$

Given a BMS channel W, for $n \in \mathbb{N}$, we define a random sequence of channels W_n, as $W_0 = W$, and

$$W_n = \begin{cases} W_{n-1}^-, & \text{w.p. } 1/2, \\ W_{n-1}^+, & \text{w.p. } 1/2. \end{cases} \tag{1.10}$$

Let $Z_n(W) = Z(W_n)$ be the random process that tracks the Bhattacharyya parameter of W_n. Then, from (1.7) and (1.8) we deduce that, for $n \geq 1$,

$$Z_n \begin{cases} \in \left[Z_{n-1}\sqrt{2 - Z_{n-1}^2},\, 2Z_{n-1} - Z_{n-1}^2\right], & \text{w.p. } 1/2, \\ = Z_{n-1}^2, & \text{w.p. } 1/2. \end{cases} \tag{1.11}$$

The process Z_n is a bounded super-martingale (see Proposition 9 of [37], or Lemma 2.5 of [42]), which captures the fact that the polarization process preserves the sum of the capacities of the synthetic channels.

Eventually, the proof of the fact that polar codes are capacity-achieving boils down to showing that, when $n \to \infty$, (see Proposition 10 of [37] and Lemma 2.6 of [42])

$$Z_n \xrightarrow{\text{a.s.}} Z_\infty = \begin{cases} 0, & \text{w.p. } C(W), \\ 1, & \text{w.p. } 1 - C(W). \end{cases} \tag{1.12}$$

Recall that (1.4) and (1.5) imply that $C(W) \approx 1$ if and only if $Z(W) \approx 0$, and $C(W) \approx 0$ if and only if $Z(W) \approx 1$. As a result, (1.12) means exactly that a fraction $C(W)$ of the polarized channels is noiseless, and the remaining fraction $1 - C(W)$ is completely noisy.

1.3.2 Transmission over Binary Memoryless Symmetric Channels

Let us now schematize how to achieve the capacity of a BMS channel with polar codes. We use a format that will be recurrent later on in Chapter 4 and 5. In what follows, given $N \in \mathbb{N}$, the set $\{1, \cdots, N\}$ is abbreviated as $[N]$ and, given a set $\mathcal{A} \subseteq [N]$, we denote by $\mathcal{A}^c$ its complement. We use $X_{i:j}$ as a shorthand for $(X_i, \cdots, X_j)$, with $i \leq j$.

Problem Statement. Consider a BMS channel W with input X and output Y. The aim is to transmit over W with a rate close to its capacity $C(W)$.

Design of the Scheme. Let $n \in \mathbb{N}$ and $N = 2^n$. Consider the $N \times N$ matrix G_N defined as follows,

$$G_N = B_N F^{\otimes n}, \qquad F = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \tag{1.13}$$

where $F^{\otimes n}$ denotes the n-th Kronecker power of F, and B_N is the permutation matrix defined in Section VII-B of [37]. The matrix B_N acts as a bit-reversal operator. This means that, if $w_{1:N} = v_{1:N} B_N$, then $w_i = v_j$, where the binary expansion of $i-1$ over n bits is obtained by reversing the order of the binary expansion of $j-1$ over n bits.

Let $X_{1:N}$ be a vector with N i.i.d. uniformly random components, i.e., X_i is a Bernoulli(1/2) random variable for $i \in [N]$, and set

$$U_{1:N} = X_{1:N} G_N. \tag{1.14}$$

The polarization procedure described in the previous section consists in multiplying by the matrix $F^{\otimes n}$, and the role of B_N consists in letting us decode the bits U_i in order, i.e., first U_1, then U_2, and so on. The output corresponding to the transmission of $X_{1:N}$ over the channel W is denoted by $Y_{1:N}$.

Define the sets

$$\begin{aligned} \mathcal{H}_{X|Y} &= \{i \in [N] \colon Z(U_i \mid U_{1:i-1}, Y_{1:N}) \geq 1 - \delta_N\}, \\ \mathcal{L}_{X|Y} &= \{i \in [N] \colon Z(U_i \mid U_{1:i-1}, Y_{1:N}) \leq \delta_N\}, \end{aligned} \tag{1.15}$$

where, given $(T, V) \sim p_{T,V}$, with T binary and V taking values in an arbitrary discrete alphabet $\mathcal{V}$, we define

$$Z(T \mid V) = 2 \sum_{v \in \mathcal{V}} \mathbb{P}_V(v) \sqrt{\mathbb{P}_{T|V}(0 \mid v) \mathbb{P}_{T|V}(1 \mid v)}. \tag{1.16}$$

Take T to be uniformly distributed and equal to the input of the channel, and V to be the corresponding channel output. Then, from (1.16) we recover (1.3). Clearly, the value of δ_N in (1.15) affects the performance of the code. At the end of this section, we discuss the choice of δ_N.

On the one hand, for $i \in \mathcal{H}_{X|Y}$, the bit U_i is approximately uniformly distributed and independent of $(U_{1:i-1}, Y_{1:N})$. This means that it cannot be decoded in a successive fashion, given the output of the channel and the previous bits. On the other hand, for $i \in \mathcal{L}_{X|Y}$, the bit U_i is approximately a deterministic function of $(U_{1:i-1}, Y_{1:N})$. This means that it can be decoded in a successive fashion, given the output of the channel and the previous bits. In other words, for $i \in \mathcal{H}_{X|Y}$, the bit U_i "sees" an almost completely noisy channel, and, for $i \in \mathcal{L}_{X|Y}$, the bit U_i "sees" an almost noiseless channel.

From the previous discussion on the polarization process, we have that, as N goes large, a fraction $1 - C(W)$ of the synthetic channels is completely noisy and a fraction $C(W)$ is noiseless. This translates into the fact that

$$\begin{aligned} \lim_{N \to \infty} \frac{1}{N} |\mathcal{H}_{X|Y}| &= 1 - C(W), \\ \lim_{N \to \infty} \frac{1}{N} |\mathcal{L}_{X|Y}| &= C(W). \end{aligned} \tag{1.17}$$

In order to construct a polar code for the channel W, we proceed as follows. The information bits are placed in the positions indexed by $\mathcal{L}_{X|Y}$, as these positions will be decodable in a successive fashion given the output. For this reason, the set $\mathcal{L}_{X|Y}$ will be usually denoted as $\mathcal{I}$. The remaining positions are frozen and their values are shared between the encoder and the decoder. Any choice of the frozen bits is as good as any other (see Section VI-B of [37]). Hence for the sake of simplicity, we can simply set these bits to 0.

The code construction requires knowing which synthetic channels are almost noiseless. This means knowing which positions are in $\mathcal{I}$ or, equivalently, for which values of i the synthetic channel $W_n^{(i)}$ has a Bhattacharyya parameter close to 0. In practice, given a block length N and a rate R, we have to find the NR synthetic channels with the smallest Bhattacharyya parameters. Once again, the BEC provides the simplest possible case. Indeed, for the transmission over an erasure channel, the Bhattacharyya parameters of all the synthetic channels can be computed in $O(N)$, as mentioned in [45]. In general, the problem is hard, as the cardinality of the output alphabet of $W_n^{(i)}$ is exponential in N. Hence, computing the exact transition probabilities of these channels seems intractable. In Arıkan's original paper [37], it is proposed to estimate these Bhattacharyya parameters by a Monte-Carlo method. In [46], Tal and Vardy show that, by performing the evaluation approximately, the construction has only linear complexity. A framework where the algorithms of [46] and new related algorithms can be analyzed and compared is provided in [47].

Furthermore, in order to communicate close to capacity, the construction of polar codes has to be tailored to the specific transmission channel. In general, given channels W and W' with $C(W) = C(W')$, a polar code designed to be capacity-achieving for W will not be capacity-achieving for W'. Several techniques for constructing universal polar codes, i.e., polar codes that can achieve the compound capacity of the whole class of BMS channels, are presented in [48, 49].

Encoding. We place the information into the positions indexed by $\mathcal{I}$, hence let $\{u_i\}_{i\in\mathcal{I}}$ denote the information bits to be transmitted. The remaining positions are filled with all 0s. As $G_N = G_N^{-1}$, the vector $x_{1:N} = u_{1:N}G_N$ is transmitted over the channel. As discussed in Section VII of [37], by exploiting the particular structure of the matrix G_N, it is possible to perform this matrix multiplication with complexity $O(N \log_2 N)$.

Decoding. The decoder receives $y_{1:N}$ and computes the estimate $\hat{u}_{1:N}$ of $u_{1:N}$ according to the rule

$$\hat{u}_i = \begin{cases} \arg\max\limits_{u\in\{0,1\}} \mathbb{P}_{U_i|U_{1:i-1},Y_{1:N}}(u \mid u_{1:i-1}, y_{1:N}), & \text{if } i \in \mathcal{I}, \\ u_i, & \text{otherwise,} \end{cases} \tag{1.18}$$

where the probabilities $\mathbb{P}_{U_i|U_{1:i-1},Y_{1:N}}(u \mid u_{1:i-1}, y_{1:N})$ can be computed recursively with complexity $O(N \log_2 N)$ (see Section VIII of [37]). This is a successive cancellation (SC) decoder, as the estimates are produced one by one with a single pass on the data (as opposed to iterative decoding).

In order to improve the performance of the successive decoder described above, several other decoding algorithms have been proposed. Optimal maximum a posteriori (MAP) decoders are implemented via the Viterbi algorithm [50] and via sphere

decoding [51], but they are practical only for relatively short block lengths. A linear programming (LP) decoder is introduced in [52], and the performance under belief-propagation decoding is considered in [53]. The stopping set analysis for the transmission over the BEC is also provided in [54].

A successive cancellation list (SCL) decoder with space complexity $O(LN)$ and time complexity $O(LN \log_2 N)$ is proposed in [55], where L is the size of the list. Empirically, the use of several concurrent decoding paths yields an error probability comparable to that under MAP decoding with practical values of the list size. In addition, by adding only a few extra bits of cyclic redundancy check (CRC) precoding, the results are comparable with the performance of current state-of-the-art LDPC codes. Motivated by these empirical observations, in Chapter 3 of this thesis, we will deal with provable bounds on the performance of list decoders.

Performance. The block error probability P_{B} can be upper bounded by the sum of the Bhattacharyya parameters of the channels that are not frozen (see Proposition 2 of [37]). In formulae,

$$P_{\mathrm{B}} \leq \sum_{i \in \mathcal{I}} Z(U_i \mid U_{1:i-1}, Y_{1:N}) \leq N\delta_N. \tag{1.19}$$

In Arıkan's original paper [37], δ_N is upper bounded by $N^{-5/4}$, hence P_{B} is $O(N^{-1/4})$. This bound is refined in [56], where it is shown that P_{B} is $O(2^{-N^{\beta}})$ for any $\beta \in (0, 1/2)$. Several new results on the finite-length performance of polar codes will be presented in Chapter 2.

1.4 Reed-Muller Codes

Unlike polar codes that were discovered by Arıkan in 2008, Reed-Muller (RM) codes are among the oldest known codes. They were introduced by Muller in 1954 [18] and, shortly thereafter, Reed proposed a majority logic decoder [19]. Given n and $v \in \mathbb{N}$, a Reed-Muller code $\mathrm{RM}(n, v)$ is a linear code of block length $N = 2^n$ and rate R, given by

$$R = \frac{\sum_{i=0}^{v} \binom{n}{i}}{2^n}.$$

It is well known that the minimum distance of this code is 2^{n-v} [12].

The relation between polar and Reed-Muller codes was first pointed out in Arıkan's seminal paper [37]. Indeed, the generator matrix of both polar and Reed-Muller codes is obtained by taking rows of the matrix $F^{\otimes n}$ defined in (1.13). However, the rule for selecting such rows is different between polar and Reed-Muller codes. Performance comparisons were carried out in [57, 58]. It was observed in Section 6.1.2 of [42] that Dumer's recursive algorithm for Reed-Muller codes [59] is similar to the successive cancellation decoder for polar codes. List decoding was used also to improve the performance of Reed-Muller codes [60, 61]. Furthermore, recursive techniques can be employed to decode nested polarized codes, in which the splitting process ends at various short Reed-Muller codes, instead of the single information bits used as end nodes in polar codes [62, 63]. A hybrid design that

combines the construction of Reed-Muller and polar codes is introduced in [64]. Numerical simulations and analytical results suggest that Reed-Muller codes have a bad performance under successive and iterative decoding, but they outperform polar codes under MAP decoding [37, 53]. In Chapter 6 of this thesis, we will have to say more about the relation between these two classes of codes.

In summary, polar codes are capacity-achieving and Reed-Muller codes seem to be even better when using an optimal decoder! Hence, it is reasonable to conjecture that Reed-Muller codes also achieve capacity. Actually, this idea appears to be rather old. Already at the end of the 1960s, it was discussed privately by Kasami, Lin, and Peterson. In 1991, at the IEEE Information Theory Workshop, Lin explicitly mentioned this possibility in his talk entitled "RM Codes Are Not So Bad". In 1994, Dumer and Farrell suggested, as an open problem, the evaluation of a quantity that equals 1 if and only if Reed-Muller codes achieve capacity on the BEC [65]. Since then, similar ideas have been discussed by several other authors [10, 66–69]. Short Reed-Muller codes with erasures were investigated in [66, 67], and it was observed numerically that the block error probability is quite close to that of random codes. Costello and Forney conjectured in [10] that the sequence of rate-1/2 self-dual Reed-Muller codes achieves capacity on the binary-input AWGN channel. The regimes in which the rate is either very high or very low were studied by Abbe, Shpilka, and Wigderson in [68, 69]: for rates approaching either 0 or 1 with sufficient speed, it is shown that Reed-Muller codes can correct almost all erasure patterns up to the capacity limit[1]; for rates approaching 0 fast enough, it is also proved that Reed-Muller codes can correct random error patterns up to the capacity limit. However, the regime that is typically of interest in coding theory requires that the rates of the codes tend to a constant $\in (0, 1)$. As promised by the title of this thesis, we will eventually solve this conjecture in Chapter 7. We will discuss generalizations and extensions in Chapter 8.

1.5 Unified Scaling

Consider the transmission of a code over a noisy channel. In a wide sense, with unified scaling, we indicate the study of the relation of the relevant parameters, i.e., rate, block length, error probability, and quality of the communication channel. Concretely, consider the plots in Figure 1.4: they represent the performance of the family of codes $\mathcal{C}$ with rate $R = 0.5$. Different curves correspond to codes of different block lengths N. The codes are transmitted over a family of channels $\mathcal{W}$ parameterized by z that is represented on the horizontal axis. On the vertical axis, we represent the block error probability P_{B}. The error probability is an increasing function of z, which means that the channel gets "better" as z decreases. The parameter z indicates the quality of the transmission channel W and, for example, it could be set to $Z(W)$ or to $1 - C(W)$. Let us assume that there exists a threshold z^* such that, if $z < z^*$, then P_{B} tends to 0 as N grows large, whereas if $z > z^*$, then P_{B} tends to 1 as N grows large. For example, if the family of codes $\mathcal{C}$ is capacity-achieving, then we can think of the threshold z^* as the channel parameter such that $C(W) = R$. In the example of Figure 1.4, we have that $z^* = 0.5$.

[1] Some effort is required to define capacity for rates approaching 0 or 1. See Definition 16 of [68] or Section 8.3 of this thesis for further details.

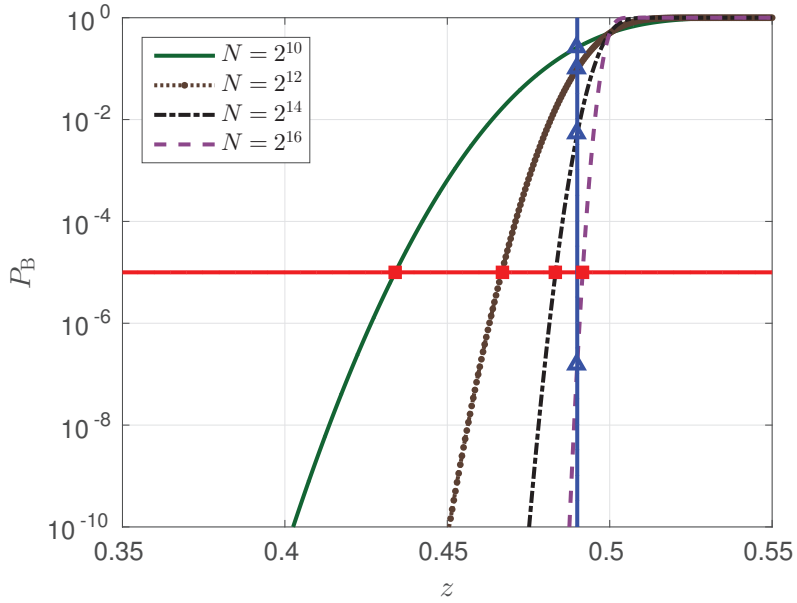


Figure 1.4 – Performance of the family of codes $\mathcal{C}$ with rate $R = 0.5$, transmitted over the family of channels $\mathcal{W}$ with threshold $z^* = 0.5$. Each curve corresponds to a code of an assigned block length N. On the x-axis, it is represented the channel parameter z and, on the y-axis, the error probability P_{B}. The error exponent regime captures the behavior of the vertical/blue cut of fixed channel parameter z (or, equivalently, of fixed gap to threshold $z^* - z$). The scaling exponent regime captures the behavior of the horizontal/red cut of fixed error probability P_{B}. The error floor regime captures the behavior of a single curve of fixed block length N.

The oldest approach for analyzing the performance of the family $\mathcal{C}$ is based on computing the *error exponent*. We pick any channel parameter $z < z^*$. Then, by definition of z^*, the error probability tends to 0 as N grows large. The error exponent regime quantifies this statement and computes how the error probability varies as a function of the block length. This approach is pictorially represented as the vertical/blue cut in Figure 1.4. The best possible scaling is obtained by considering random codes that give

$$P_{\mathrm{B}} = e^{-NE(R,W)+o(N)},$$

where $E(R, W)$ is the so-called error exponent [7]. For a fairly recent survey on how to determine the error exponent for various random ensembles, see [70].

Another approach is based on computing the *scaling exponent*. We pick a target error probability P_{B}. Then, by definition of z^*, the gap between the threshold and the channel parameter $z^* - z$ tends to 0 as N grows large. The scaling exponent regime quantifies this statement and computes how the gap to the threshold varies as a function of the block length. This approach is pictorially represented as the horizontal/red cut in Figure 1.4. From a practical viewpoint, we are interested in such a regime, as we typically have a certain requirement on the error probability and look for the shortest code possible. For specific classes of codes, this approach was put forward in [71,72]. As a benchmark, a sequence of papers that started in the

early 1960s with [73, 74] and recently culminated in [75, 76] shows that the smallest possible block length N required to achieve a gap $z^* - z$ to the threshold with a fixed error probability P_{B} is such that

$$N \approx \frac{V(Q^{-1}(P_{\mathrm{B}}))^2}{(z^* - z)^2}, \tag{1.20}$$

where $Q(\cdot)$ is the tail probability of the standard normal distribution, i.e.,

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{+\infty} \exp\left(-u^2/2\right) du, \tag{1.21}$$

and V is referred to as channel dispersion and measures the stochastic variability of the channel relative to a deterministic channel with the same capacity. In general, if N is $\Theta\left(1/(z^* - z)^{\mu}\right)$, then we say that the family of codes $\mathcal{C}$ has scaling exponent μ. Hence, by (1.20), the most favorable scaling exponent is $\mu = 2$ and it is achieved by random codes. Furthermore, for a large class of ensembles of LDPC codes and channel models, the scaling exponent is also $\mu = 2$ [77]. However, it has to be pointed out that the threshold of such LDPC ensembles does not converge to capacity.

In summary, in the error exponent regime, we compute how fast P_{B} goes to 0 as a function of N when $z^* - z$ is fixed; and in the scaling exponent regime, we compute how fast $z^* - z$ goes to 0 as a function of N when P_{B} is fixed. Then, a natural question is to ask how fast do *both* P_{B} *and* $z^* - z$ go to 0 as functions of N. In other words, we can describe a trade-off between the speed of decay of the error probability and the speed of decay of the gap to capacity as functions of the block length. This intermediate approach is named the *moderate deviations* regime and is studied for random codes in [78].

The last scaling approach we consider concerns the so-called *error floor* regime. We pick a code of assigned block length N and rate R. Then, we compute how the error probability P_{B} behaves as a function of the channel parameter z. This corresponds to taking into account one of the four curves in Figure 1.4. This is a notion that became important when iterative coding schemes were introduced. For such schemes, it was observed that frequently the individual curves $P_{\mathrm{B}}(z)$ show an abrupt change of slope, from very steep to very shallow, when going from bad channels to good channels (see, e.g., Figure 1.5). The region where the slope is very shallow was dubbed the error floor region. More specifically, if we consider a parallel concatenated turbo code, then there is a fixed number of low-weight codewords, regardless of the block length N (see Section 6.9 of [44]). The same behavior can be observed for the ensemble average of LDPC codes, when the minimal variable-node degree $\mathtt{l}_{\min}$ is equal to 2. This means that, in the error floor region, the block error probability is dominated by a term that is independent of N and scales as z^w, where w denotes the minimal weight of a non-zero codeword. If the minimal variable-node degree $\mathtt{l}_{\min}$ is at least 3, then the number of low-weight codewords vanishes with N and the error probability scales as $z^w/N^{w(\mathtt{l}_{\min}/2-1)}$. For a more precise statement, see Theorem D.32 in Appendix D of [44].

1.6 Coding for Non-standard Channels

A non-standard communication scenario is represented by any model that differs from the simple point-to-point binary memoryless symmetric setting: the channel

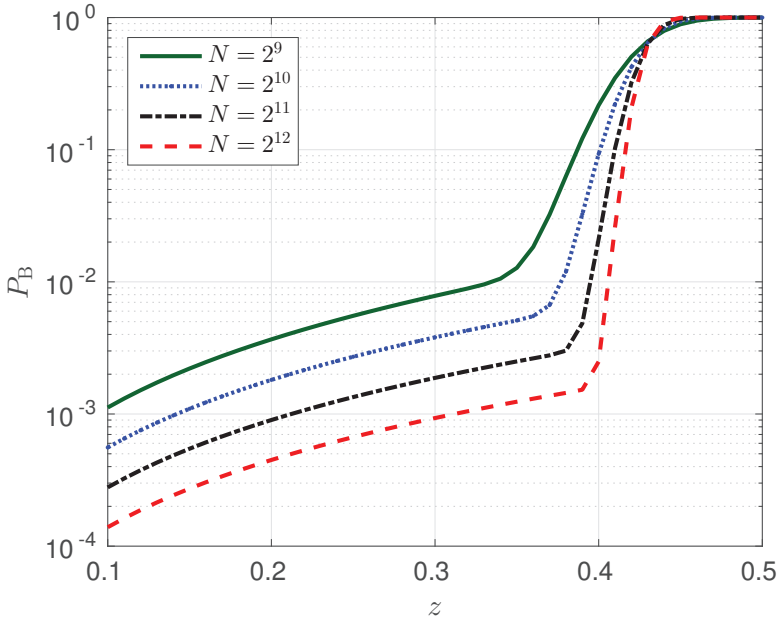


Figure 1.5 – Performance of the family of $(3, 6)$-regular LDPC codes transmitted over the BEC with erasure probability z. The waterfall region in which the error probability decreases sharply is clearly distinguishable from the error floor region in which the decay is much slower.

might receive q-ary inputs, for $q \in \mathbb{N}$, or real-valued inputs, it might be asymmetric, it might have memory, and there might be several users that want to communicate.

Concerning non-binary models, polar coding schemes were quickly generalized to arbitrary memoryless symmetric channels. Channel polarization for q-ary input alphabets is first discussed in [79] and more general constructions based on arbitrary kernels are described in [80]. In general, various algebraic structures on the input alphabet were exploited to build polar codes [81–85]. Polar coding schemes that achieve the capacity of the AWGN channel are developed in [86]. On the contrary, the capacity-achieving nature of SC-LDPC codes was proved only for binary-input channels, and it remains an open problem for the transmission over channels with a non-binary input alphabet. The iterative decoding threshold on the BEC for non-binary SC-LDPC code ensembles is investigated in [87,88], and the corresponding threshold saturation is proved in [89]. The threshold analysis under windowed decoding is provided in [90].

Furthermore, the original point-to-point polar coding scheme was extended to lossless and lossy source coding [91–93] and to many multi-terminal scenarios. Examples include Gelfand-Pinsker, Wyner-Ziv, and Slepian-Wolf problems [42, 94, 95], multiple-access channels [81, 84, 96–98], broadcast channels [99, 100], interference channels [101, 102], relay channels [103–106], wiretap channels [103, 107–110], write once memories [111], arbitrarily permuted parallel channels [112], and multiple description coding [113, 114]. Coding solutions for multi-user scenarios based on polar codes and on commercial off-the-shelf codes are also provided in [115]. The fundamental purpose of this line of work is to implement in a low-complexity fashion the

classic random coding schemes in network information theory, for example, binning, rate-splitting, superposition coding, decode-forward relaying, compress-forward relaying, and so on. We will focus more specifically on broadcast and asymmetric channels in Chapter 4 and 5, respectively.

1.7 Outline and Contributions of this Thesis

Before digging into the details, let us give a bird's eye view on the structure of this thesis. The topic of the first two chapters is *scaling*: in Chapter 2, we provide a unified framework for the finite-length analysis of polar codes under successive cancellation decoding, and, in Chapter 3, we study the scaling exponent of list decoders. The topic of the next two chapters is coding for *non-standard* communication scenarios: in Chapter 4, we deal with broadcast channels and, in Chapter 5, with asymmetric channels. Chapter 6 serves as an interlude to move from more recent coding techniques, primarily polar codes, towards older and more structured ones, primarily Reed-Muller codes. The topic of the following two chapters is how to achieve *capacity via symmetry*: in Chapter 7, we settle a decade-long conjecture by showing that any sufficiently symmetric family of codes achieves capacity on the binary erasure channel and, in Chapter 8, we provide several generalizations of this result. Eventually, in Chapter 9, we summarize the main contributions of our work and outline future research directions.

This thesis tells a story that starts with polar codes and ends with Reed-Muller codes. However, the technical material of each chapter is intended to be self-contained. A notable exception to this principle is Chapter 8; it contains some generalizations of the main result of Chapter 7 and, for this reason, should be read afterwards.

Unified Scaling of Polar Codes

Our first contribution consists in the unified characterization of the finite-length performance of polar codes. More specifically, we consider the transmission of a polar code of block length N and rate R over a BMS channel W with capacity $C(W)$ and Bhattacharyya parameter $Z(W)$, and we let P_{B} be the error probability under successive cancellation decoding. In previous work, two main regimes were studied. In the error exponent regime, the channel W and the rate $R < C(W)$ are fixed, and it was proved that the error probability P_{B} scales roughly as $2^{-\sqrt{N}}$. In the scaling exponent regime, the channel W and the error probability P_{B} are fixed, and it was proved that the gap to capacity $C(W) - R$ scales as $N^{-1/\mu}$. Here, μ is the scaling exponent and it depends on the channel W. A heuristic computation for the BEC gives $\mu = 3.627$, and it was shown that, for any BMS channel, $3.579 \leq \mu \leq 5.702$.

In **Chapter 2**, we develop a *unified framework* to characterize the relationship of R, N, P_{B}, and the quality of the channel W. First, we provide the *tighter upper bound* $\mu \leq 4.714$, valid for any BMS W. With the same technique, we obtain the upper bound $\mu \leq 3.639$ for the case of the BEC; this upper bound approaches very closely the heuristically derived value for the scaling exponent of the erasure channel. Second, we consider a *moderate deviations* regime and we study how fast both the gap to capacity $C(W) - R$ and the error probability P_{B} simultaneously go to 0 as

N goes large. Third, we prove that polar codes are not affected by *error floors*. To do so, we fix a polar code of block length N and rate R, we let the channel W vary, and we show that P_{B} scales roughly as $Z(W)^{\sqrt{N}}$.

Scaling Exponent of List Decoding

Compared to random codes that have a scaling exponent of 2, the tight bounds discussed above enable us to conclude that polar codes possess a less favorable trade-off between block length and gap to capacity under successive cancellation decoding. Therefore, the most natural question that comes to mind is how to improve such scaling exponent. Motivated by the significant performance gains that polar codes experience under successive cancellation *list* decoding [55], we study the scaling exponent of list decoding as a function of the list size.

In **Chapter 3**, we prove that, by adding a *list of finite size* to the MAP decoder, the *scaling exponent stays unaffected* for any BMS channel and for any sequence of linear codes such that their minimum distance is unbounded as $N \to \infty$. To do so, we develop a *Divide and Intersect* (DI) procedure, in order to lower bound the error probability under MAP decoding with list size L. In particular, the result applies to polar codes, as their minimum distance tends to infinity as $N \to \infty$.

The DI technique is rather general. Indeed, when the transmission takes place over the BEC, we prove a similar result for genie-aided SC decoding: the *scaling exponent* remains *constant* for any *fixed number of helps from the genie*.

How to Achieve Marton's Region for Broadcast Channels

With **Chapter 4**, we move to the second main topic of this thesis, i.e., coding techniques for non-standard channels. In particular, we consider the two-user discrete memoryless broadcast channel (DM-BC) and we present a *polar coding scheme* that *achieves Marton's region* with both common and private messages. This is the *best achievable rate region* known to date, and it is tight for all classes of two-user DM-BCs whose capacity regions are known. To accomplish this task, we first construct polar codes for both the superposition, as well as the binning strategy. By combining these two schemes, we obtain Marton's region with private messages only. Finally, we show how to handle the case of common information. The proposed coding schemes possess the usual advantages of polar codes, i.e., low encoding complexity, low decoding complexity, and super-polynomial decay rate of the error probability.

We follow the lead of Goela, Abbe, and Gastpar, who recently introduced polar codes emulating the superposition and binning schemes [100]. In order to align the polar indices, for both schemes, their solution involves some degradation constraints that are assumed to hold between the auxiliary random variables and the channel outputs. To remove these constraints, we guarantee the proper alignment of the polarized indices by means of a *chaining construction*. This technique was originally introduced in [48] to construct universal codes and in [110] to achieve strong security guarantees on degraded wiretap channels. The idea is as simple as it is powerful: by transmitting several code blocks and by repeating suitable parts of the information bits in neighboring blocks, we construct polar codes that are good for the transmission over more than one channel at the same time. Because of its generality, the chaining construction has been applied by different authors to various other multi-

terminal scenarios [116–119]. Another non-standard setting in which this technique proves useful is the transmission over asymmetric channels, as we will see in the chapter that immediately follows.

How to Achieve the Capacity of Asymmetric Channels

In the previous chapter, we have described low-complexity polar coding schemes for the broadcast channel. In **Chapter 5**, we consider the transmission over an *asymmetric* channels and survey capacity-achieving coding techniques. In particular, we take the point of view of modern coding theory and discuss how recent advances in coding for symmetric channels help provide more efficient solutions for the asymmetric case. We consider, in more detail, three basic coding paradigms.

The first one is *Gallager's scheme* [120] that consists of concatenating a linear code with a non-linear mapper so that the input distribution can be appropriately shaped. We explicitly show that both polar codes and spatially coupled codes can be employed in this scenario. Furthermore, we derive a scaling law between the gap to capacity, the cardinality of the input and output alphabets, and the required size of the mapper.

The second one is an *integrated approach* in which the coding scheme is used *both* for source coding, in order to create codewords distributed according to the capacity-achieving input distribution, *and* for channel coding, in order to provide error protection. Such a technique has been recently introduced by Honda and Yamamoto in the context of polar codes [121], and we show how to apply it also to the design of sparse graph codes.

The third paradigm is based on a *chaining construction* similar to the one introduced in the previous chapter. The idea is to separate the two tasks of source coding and channel coding by chaining together several codewords, and it is originally due to Böcherer and Mathar [122]. Here, we show that this technique yields provably capacity-achieving coding schemes. In particular, we present conditions for the source code and the channel code, and we describe how to combine *any* source code with *any* channel code that fulfill those conditions, in order to achieve the capacity of asymmetric channels. Furthermore, we prove that polar codes, spatially coupled codes, and arithmetic codes are suitable as basic building blocks of the proposed approach.

Rather than focusing on the exact details of the schemes, the purpose of this chapter is to present different coding strategies that can then be implemented with many variants. There is no absolute winner and, in order to understand the most suitable technique for a specific application scenario, we provide a detailed comparison that takes into account several performance metrics.

Interlude – from Polar to Reed-Muller Codes

Chapter 6 marks the transition towards the third and last topic of this thesis, as it ideally connects the families of polar and Reed-Muller codes. In particular, we explore the relationship between these two coding techniques and, by doing so, we present a *new coding scheme* that *significantly improves* upon the *performance of polar codes* at practical block lengths.

Our starting point is the experimental observation that Reed-Muller codes have a smaller error probability than polar codes under MAP decoding. This motivates us to introduce the family of codes $\{\mathcal{C}_\alpha\}$, for $\alpha \in [0,1]$. Such a family "interpolates" between Reed-Muller and polar codes, in the sense that $\mathcal{C}_\alpha\big|_{\alpha=1}$ is the original polar code, and $\mathcal{C}_\alpha\big|_{\alpha=0}$ is a Reed-Muller code. Using numerical observations, we remark that the error probability under MAP decoding is an increasing function of α. As MAP decoding has in general exponential complexity, we also consider practical decoding schemes, such as the belief-propagation or the successive cancellation list decoder. The result is that, even under these low-complexity decoding algorithms, the performance of polar codes is boosted by moving along the family $\{\mathcal{C}_\alpha\}$. We demonstrate this performance gain via numerical simulations for the transmission over the BEC and the binary-input AWGN channel. Such a gain could be substantial in the sense of the reduction of the scaling exponent: according to numerical simulations performed for $N = 2^{10}$ over the BEC, the error probability under MAP decoding for the transmission of $\mathcal{C}_\alpha$ for α sufficiently small is very close to that of random codes that have the best possible scaling exponent. As a result, the use of codes from the family $\{\mathcal{C}_\alpha\}$ potentially improves the speed at which capacity is reached.

Capacity via Symmetry I: A Proven Conjecture

Eventually, in **Chapter 7**, we give the proof of the conjecture that has been promised since the very title of this thesis: *Reed-Muller codes achieve capacity for the transmission over the BEC under MAP decoding*. Actually, we prove a much more general result: *any sequence of linear codes with doubly transitive permutation group achieves capacity* for the transmission over the BEC under MAP decoding. In other words, we show that symmetry alone implies asymptotically optimal performance.

The proof exploits three main ingredients coming from different areas of information theory and computer science:

- the *code symmetry*, coming from algebraic coding theory;
- the *sharp threshold* framework applied to the measure of monotone symmetric sets, a very powerful and popular tool in theoretical computer science;
- the *area theorem* for extrinsic information transfer functions, coming from iterative coding theory.

Capacity via Symmetry II: Generalizations

In the previous chapter, we have proved one basic, yet fundamental, result: codes with sufficient symmetry achieve capacity on the BEC under bit-MAP decoding for any rate in $(0,1)$. In **Chapter 8**, we provide some generalizations: we consider the limiting regimes in which the rate of the code is either very low or very high, and we show how to strengthen the results regarding the bit-MAP threshold to the block-MAP threshold.

It has been recently proved that Reed-Muller codes can correct almost all erasure patterns up to the capacity limit for rates approaching either 0 or 1 with a specific speed [68, 69]. By exploiting the proof technique developed in the previous chapter,

we show that Reed-Muller codes achieve capacity in another non-overlapping regime, i.e., for a different speed of convergence of the rate.

For the comparison between bit-MAP and block-MAP thresholds, let us point out that the result in the previous chapter holds under bit-MAP decoding. Thus, a natural question is what happens under block-MAP decoding. By exploiting further symmetries of the code, it is possible to show that the bit-MAP threshold is sharp enough that the block erasure probability also converges to 0. However, this technique relies heavily on the fact that the transmission is over an erasure channel.

Our main technical contribution in this chapter consists in presenting a more general approach to passing from the bit-MAP error probability to the block-MAP error probability. This approach is based on the careful analysis of the weight distribution of Reed-Muller codes. In particular, our result has the following flavor: assume that the bit-MAP error probability for the transmission over any BMS channel decays as $N^{-\delta}$, for some $\delta > 0$; then, the block-MAP error probability also converges to 0. Let us highlight that the proposed technique does not apply only to the special case of the erasure channel, but it is valid for the transmission over any binary memoryless symmetric channel. Hence, this result can be thought of as a first step in extending the proof that Reed-Muller codes are capacity-achieving to the general case.

Conclusions and Perspectives

In **Chapter 9**, we summarize the *main contributions* of this thesis, discuss *open questions*, and describe how the *novel technical tools* developed so far have already proved useful for other problems. We conclude by presenting three fairly wide research directions, one for each of the main topics considered in this thesis:

1. boost the performance of polar codes at practical block lengths, by devising a coding scheme with a provably better scaling exponent;

2. consider a multi-user setting, and transmit with a low-complexity technique in a rate region that was not previously known to be information-theoretically achievable;

3. find a low-complexity algorithm to decode Reed-Muller codes with a performance close to that of the MAP decoder.

2 Unified Scaling of Polar Codes

C'è davvero bisogno di domande retoriche?

Do you really need rhetorical questions?

In this chapter[1], we provide a unified view on the performance analysis of polar codes and present several results about the scaling of the parameters of interest, namely, the rate R, the block length N, the block error probability under SC decoding P_{B}, and the quality of the channel W.

In Section 2.1, we review the existing literature on finite-length scaling of polar codes. In Section 2.2, we summarize our main contributions and, in the following three sections, we describe them in detail: in Section 2.3, we present the new upper bound on the scaling exponent; in Section 2.4, we address the moderate deviations regime; and in Section 2.5, we show that polar codes are not affected by error floors. We defer some of the proofs to the appendix in Section 2.6.

2.1 Related Work

Since the introduction of polar codes in the seminal paper [37], their performance has been extensively studied in different regimes. The error exponent, scaling exponent, moderate deviations, and error floor regimes have been described in Section 1.5. However, we will quickly recall their definitions when presenting the existing results for polar codes.

In the *error exponent* regime, the rate $R < C(W)$ is fixed, and it is studied how the error probability P_{B} scales as a function of the block length N. As pointed out at the end of Section 1.3.2, in [56] it is proved that the block error probability under SC decoding behaves roughly as $2^{-\sqrt{N}}$. This result is further refined in [125], where

[1] The material of this chapter is based on joint work with S. H. Hassani and R. Urbanke [123,124].

it is shown that $\log_2(-\log_2 P_{\mathrm{B}})$ scales as

$$\frac{\log_2 N}{2} + \frac{\sqrt{\log_2 N}}{2} \cdot Q^{-1}\left(\frac{R}{C(W)}\right) + o(\sqrt{\log_2 N}), \tag{2.1}$$

where $Q(\cdot)$ is the tail probability of the standard normal distribution defined in (1.21). This last result holds both under SC decoding and under optimal MAP decoding.

In the *scaling exponent* regime, the error probability P_{B} is fixed, and it is studied how the gap to capacity $C(W)-R$ scales as a function of the block length N. In [126], the scaling exponent is defined as the value of μ such that

$$\lim_{N\to\infty, N^{1/\mu}(C(W)-R)=z} P_{\mathrm{B}}(N,R,W) = f(z), \tag{2.2}$$

for some function $f(z)$, that is called *mother curve*.

It is an open question to prove that the limit (2.2) exists. Note that the value of μ depends on the particular channel taken into account. The authors of [126] provide a heuristic method for computing the scaling exponent for the transmission over the BEC under SC decoding; this method yields $\mu \approx 3.627$. Furthermore, in [127] it is shown that the block length scales polynomially fast with the inverse of the gap to capacity, while the error probability is upper bounded by $2^{-N^{0.49}}$. Universal bounds on μ, valid for any BMS channel under SC decoding, are presented in [128]: the scaling exponent is lower bounded by 3.579 and upper bounded by 6. In addition, it is conjectured that the lower bound on μ can be increased up to 3.627, i.e., up to the value heuristically computed for the BEC. The upper bound on μ is further refined to 5.702 in [129].

In the *moderate deviations* regime, neither the rate nor the error probability are fixed, but it is studied how the gap to capacity $C(W) - R$ and the error probability P_{B} jointly scale as functions of the block length N. There is no prior work on this regime for polar codes.

In the *error floor* regime, the code is fixed, i.e., the rate R and block length N are fixed, and it is studied how the error probability P_{B} scales as a function of the channel parameter. In [54] it is proved that the stopping distance of polar codes scales as $\sqrt{N}$, which implies good error floor performance under BP decoding. The authors of [54] also provide simulation results that show no sign of error floor for transmission over the BEC and over the binary-input AWGN channel.

2.2 Main Results

Our contributions in this chapter address the *scaling exponent*, the *moderate deviations*, and the *error floor* regimes, and they can be summarized as follows.

New universal upper bound on scaling exponent. We show that $\mu \leq 4.714$ for any BMS channel and that $\mu \leq 3.639$ for the BEC. Basically, this result *improves by* 1 the *previous upper bound* valid for any BMS channel and *approaches* closely the *value* 3.627 that has been *heuristically computed* for the BEC. The proof technique consists in relating the scaling exponent to the supremum of some function and, then, in describing an interpolation algorithm

to obtain a provable upper bound on this supremum. The values 4.714 for any BMS channel and 3.639 for the BEC are obtained for a particular number of samples used by the algorithm and they can be slightly improved simply by running the algorithm with a larger number of samples.

Moderate deviations. We *unify* the two perspectives of the *error exponent* and the *scaling exponent* by letting both the gap to capacity $C(W) - R$ and the error probability P_{B} go to 0 as functions of the block length N. In particular, we describe a trade-off between the speed of decay of P_{B} and the speed of decay of $C(W) - R$. In the limit in which the gap to capacity is arbitrarily small but independent of N, this trade-off recovers the result of [56], where it is shown that P_{B} scales roughly as $2^{-\sqrt{N}}$.

Absence of error floors. We prove that *polar codes* are *not affected by error floors*. To do so, we consider a polar code of block length N and rate R designed for the transmission over a channel W'. Then, we look at the performance of this fixed code over other channels W that are "better" than W'; and we study the error probability P_{B} as a function of the Bhattacharyya parameter $Z(W)$. Note that the code is fixed and the channel varies, which means that we do not choose the optimal polar indices for W. In particular, we prove that P_{B} scales roughly as $Z(W)^{\sqrt{N}}$, and this result is in agreement with the error exponent regime.

2.3 New Universal Upper Bound on Scaling Exponent

In this section, we propose an improved upper bound on the scaling exponent that is valid for the transmission over any BMS channel W. First of all, we relate the value of the scaling exponent μ to the supremum of some function. Secondly, we provide a provable bound on this supremum, which gives us a provably valid choice for μ, i.e., $\mu = 4.714$ for any BMS channel and $\mu = 3.639$ for the BEC.

2.3.1 Statement and Discussion

Theorem 2.1 (From Eigenfunction to Scaling Exponent)**.** *Assume that there exists a function $h(x) : [0,1] \to [0,1]$ such that $h(0) = h(1) = 0$, $h(x) > 0$ for any $x \in (0,1)$, and, for some $\mu > 2$,*

$$\sup_{x \in (0,1), y \in [x\sqrt{2-x^2}, 2x-x^2]} \frac{h(x^2) + h(y)}{2h(x)} < 2^{-1/\mu}. \tag{2.3}$$

Consider the transmission over a BMS channel W with capacity $C(W)$ by using a polar code of rate $R < C(W)$. Fix $p_{\mathrm{B}} \in (0,1)$ and assume that the block error probability under successive cancellation decoding is at most *p_{B}. Then, it suffices to have a block length N such that*

$$N \leq \frac{\beta_1}{(C(W) - R)^{\mu}}, \tag{2.4}$$

where β_1 is a universal constant that does not depend on W, but only on p_{B}. If W is a BEC, a less stringent hypothesis on μ is required for (2.4) *to hold. In particular,*

the condition (2.3) *is replaced by*

$$\sup_{x\in(0,1)} \frac{h(x^2)+h(2x-x^2)}{2h(x)} < 2^{-1/\mu}. \tag{2.5}$$

Theorem 2.2 (Valid Choice for Scaling Exponent)**.** *Consider the transmission over a BMS channel W with capacity $C(W)$ by using a polar code of rate $R < C(W)$. Fix $p_{\mathrm{B}} \in (0,1)$ and assume that the block error probability under successive cancellation decoding is* at most *p_{B}. Then, it suffices to have a block length N upper bounded by* (2.4) *with $\mu = 4.714$. Furthermore, if W is a BEC, then* (2.4) *holds with $\mu = 3.639$.*

Before proceeding with the proofs, it is useful to discuss two points. The first remark focuses on the role of the function $h(x)$ and heuristically explains why the value of the scaling exponent is linked to the existence of a function that fulfills condition (2.3) (condition (2.5) for the BEC). Note that the remark contains just a heuristic discussion and the proofs of Theorems 2.1 and 2.2 do not depend on this explanation. Hence, we will not be concerned with mathematical rigor and, in particular, with the existence of the discussed eigenfunctions/eigenvalues. The second remark points out that we can let the error probability tend to 0 polynomially fast in N and maintain the same scaling between gap to capacity and block length.

Remark 2.1 (Heuristic Interpretation of Function $h(x)$)**.** *First, let W be a BEC and consider the linear operator T_{BEC} defined as*

$$T_{\mathrm{BEC}}(g) = \frac{g(x^2)+g(2x-x^2)}{2}, \tag{2.6}$$

where $g(x)$ is a bounded and real valued function over $[0,1]$. The relation between the Bhattacharyya process Z_n and the operator T_{BEC} is given by

$$\mathbb{E}\left[g(Z_n) \mid Z_0 = x\right] = \overbrace{T_{\mathrm{BEC}} \circ T_{\mathrm{BEC}} \circ \cdots \circ T_{\mathrm{BEC}}(g)}^{n \text{ times}} = T^n_{\mathrm{BEC}}(g), \tag{2.7}$$

where the formula comes from a straightforward application of (1.2)*. A detailed explanation of the dynamics of the functions $T^n_{\mathrm{BEC}}(g)$ is provided in Section III of [128]. In short, a simple check shows that $\lambda = 1$ is an eigenvalue of the operator T_{BEC} with eigenfunctions $v_0(x) = 1$ and $v_1(x) = x$. Let λ^* be the largest eigenvalue of T_{BEC} other than $\lambda = 1$ and define μ^* as $\mu^* = -1/\log_2 \lambda^*$. Then, the heuristic discussion of [128] leads to the fact that μ^* is the largest candidate that we could plug in* (2.5)*. For this choice, the function $h(x)$ represents the eigenfunction associated with the eigenvalue λ^*, namely,*

$$\frac{h(x^2)+h(2x-x^2)}{2} = 2^{-1/\mu^*} h(x). \tag{2.8}$$

A numerical method for the calculation of this second eigenvalue was originally proposed in [126] and yields $\mu^* = 3.627$. Furthermore, in Section III of [128] it is also heuristically explained how $\mu^* = 3.627$ gives a lower bound to the scaling exponent of the BEC.

Now, let W be a BMS channel and consider the operator T_{BMSC} defined as

$$T_{\mathrm{BMSC}}(g) = \sup_{y\in[x\sqrt{2-x^2},\,2x-x^2]} \frac{g(x^2)+g(y)}{2}. \tag{2.9}$$

Note that, differently from T_{BEC}*, the operator* T_{BMSC} *is not linear, as it involves taking a supremum. The relation between the Bhattacharyya process* Z_n *and the operator* T_{BMSC} *is given by*

$$\mathbb{E}\left[g(Z_n) \mid Z_0 = x\right] \leq T^n_{\mathrm{BMSC}}(g), \tag{2.10}$$

where the formula comes from a straightforward application of (1.11)*. Similarly, as in the case of the BEC,* $\lambda = 1$ *is an eigenvalue of* T_{BMSC} *and we write the largest eigenvalue other than* $\lambda = 1$ *as* $2^{-1/\mu^*}$*. Then, the idea is that* μ^* *is the largest candidate that we could plug in* (2.3)*, and, for this choice, the function* $h(x)$ *represents the eigenfunction associated with the eigenvalue* $2^{-1/\mu^*}$*, namely,*

$$\sup_{y \in [x\sqrt{2-x^2}, 2x-x^2]} \frac{h(x^2)+h(y)}{2} = 2^{-1/\mu^*} h(x). \tag{2.11}$$

In Section IV of [128], it is proved that the scaling exponent μ *is upper bounded by* 6*. This result is obtained by showing that the eigenvalue is at least* $2^{-1/5}$*, i.e.,* $\mu^* \leq 5$*, and that* $\mu^* + 1$ *is an upper bound on the scaling exponent* μ*. Furthermore, it is conjectured that* μ^* *is a tighter upper bound on the scaling exponent* μ*. In [129], a more refined computation of* μ^* *is presented, which yields* $\mu^* \leq 4.702$*, hence* $\mu \leq 5.702$*. In this chapter, we solve the conjecture of [128] by proving that, indeed,* μ^* *is an upper bound on the scaling exponent* μ*. In addition, we show an algorithm that guarantees a* provable *bound on the eigenvalue, thus obtaining* $\mu \leq 4.714$ *for any BMS channel and* $\mu \leq 3.639$ *for the BEC. We finally note from* (2.10) *that* T_{BMSC} *provides only an upper bound on the (expected) evolution of* Z_n*. As a result, although* $\mu \leq 4.714$ *holds universally for any channel, this bound is certainly not tight if we consider a specific BMS channel.*

Remark 2.2 (Polynomial Decay of P_{B})**.** *With some more work, it is possible to prove the following generalization of Theorem 2.1. Assume that there exists* $h(x)$ *as in Theorem 2.1 and consider the transmission over a BMS channel* W *with capacity* $C(W)$ *by using a polar code of rate* $R < C(W)$*. Then, for any* $\nu > 0$*, the block length* N *and the block error probability under successive cancellation decoding* P_{B} *are such that*

$$\begin{aligned} P_{\mathrm{B}} &\leq \frac{1}{N^{\nu}}, \\ N &\leq \frac{\beta_2}{(C(W)-R)^{\mu}}, \end{aligned} \tag{2.12}$$

where β_2 *is a universal constant that does not depend on the channel* W*. A sketch of the proof of this statement is given at the end of Section 2.3.2. The result* (2.12) *is a generalization of Theorem 2.1 in the sense that, instead of being an assigned constant, the error probability goes to* 0 *polynomially fast in* $1/N$*, and the scaling between block length and gap to capacity, i.e., the value of* μ*, stays the same. On the contrary, as described in Section 2.4, if the error probability is* $O(2^{-N^{\beta}})$ *for some* $\beta \in (0, 1/2)$*, then the scaling between block length and gap to capacity changes and depends on the exponent* β*.*

2.3.2 From Eigenfunction to Scaling Exponent

The proof of Theorem 2.1 relies on the following two auxiliary results: Lemma 2.1, proven in Appendix 2.6.1, relates the number of synthetic channels with a small enough Bhattacharyya parameter to an expected value over the Bhattacharyya process; and Lemma 2.2, proven in Appendix 2.6.2, relates the expected value over the Bhattacharyya process to the function $h(x)$.

Lemma 2.1 (From Expectation to Scaling Exponent)**.** *Let $Z_n(W)$ be the Bhattacharyya process associated with the channel W and defined in* (1.11)*. Pick any $\alpha \in (0,1)$ and assume that, for $n \geq 1$ and for some $\rho \leq 1/2$,*

$$\mathbb{E}\left[(Z_n(1-Z_n))^{\alpha}\right] \leq c_1 2^{-n\rho}, \tag{2.13}$$

where c_1 is a constant that does not depend on n. Then,

$$\mathbb{P}\left(Z_n \leq p_{\mathrm{B}}\, 2^{-n}\right) \geq C(W) - c_2\, 2^{-n(\rho-\alpha)}, \tag{2.14}$$

where $c_2 = \sqrt{2p_{\mathrm{B}}} + 2\,c_1\, p_{\mathrm{B}}^{-\alpha}$.

Lemma 2.2 (From Eigenfunction to Expectation)**.** *Let $h(x) : [0,1] \to [0,1]$ such that $h(0) = h(1) = 0$, $h(x) > 0$ for any $x \in (0,1)$, and*

$$\sup_{x\in(0,1),\, y\in[x\sqrt{2-x^2},\, 2x-x^2]} \frac{h(x^2)+h(y)}{2h(x)} \leq 2^{-\rho_1}, \tag{2.15}$$

for some $\rho_1 \leq 1/2$. Let $Z_n(W)$ be the Bhattacharyya process associated with the channel W and defined in (1.11)*. Pick any $\alpha \in (0,1)$. Then, for any $\delta \in (0,1)$, and for $n \in \mathbb{N}$,*

$$\mathbb{E}\left[(Z_n(1-Z_n))^{\alpha}\right] \leq \frac{1}{\delta}\left(2^{-\rho_1} + \sqrt{2}\frac{\delta}{1-\delta}c_3\right)^n, \tag{2.16}$$

with c_3 defined as

$$c_3 = \sup_{x\in(\epsilon_1(\alpha),\, 1-\epsilon_2(\alpha))} \frac{(x(1-x))^{\alpha}}{h(x)}, \tag{2.17}$$

where $\epsilon_1(\alpha)$, $\epsilon_2(\alpha)$ denote the only two solutions in $[0,1]$ of the equation

$$\frac{1}{2}\left(\big(x(1+x)\big)^{\alpha} + \big((2-x)(1-x)^{1/3}\big)^{\alpha}\right) = 2^{-\rho_1}. \tag{2.18}$$

If W is a BEC, a less stringent hypothesis on ρ_1 is required for (2.16) *to hold. In particular, the condition* (2.15) *is replaced by*

$$\sup_{x\in(0,1)} \frac{h(x^2)+h(2x-x^2)}{2h(x)} \leq 2^{-\rho_1}. \tag{2.19}$$

At this point, we are ready to put everything together and prove Theorem 2.1.

Proof of Theorem 2.1. Let us define

$$\rho_1 = \min\left(\frac{1}{2}, -\log_2 \sup_{x\in(0,1),\, y\in[x\sqrt{2-x^2},\, 2x-x^2]} \frac{h(x^2)+h(y)}{2h(x)}\right), \tag{2.20}$$

where $h(x)$ is the function of the hypothesis.

Set

$$\alpha = \log_2 \left(1 + \frac{2^{-1/\mu} - 2^{-\rho_1}}{2^{-1/\mu} + 2^{-\rho_1}} \right). \tag{2.21}$$

By using (2.3) and the fact that $\mu > 2$, we immediately realize that $2^{-1/\mu} - 2^{-\rho_1} > 0$, hence that $\alpha > 0$. In addition, it is easy to check that $\alpha < 1$.

Set

$$\delta = \frac{2^{-1/\mu} - 2^{-\rho_1}}{2\sqrt{2}c_3 + 2^{-1/\mu} - 2^{-\rho_1}}, \tag{2.22}$$

where c_3 is defined in (2.17). As $2^{-1/\mu} - 2^{-\rho_1} > 0$, we have that $\delta \in (0, 1)$.

In addition, $\rho_1 \leq 1/2$ and the condition (2.15) clearly follows from the definition (2.20). Consequently, we can apply Lemma 2.2, which yields formula (2.16).

Set

$$\rho = -\log_2 \left(2^{-\rho_1} + \sqrt{2} \frac{\delta}{1-\delta} c_3 \right). \tag{2.23}$$

Then, $\rho \leq \rho_1 \leq 1/2$, and we can apply Lemma 2.1 with $c_1 = 1/\delta$, which yields

$$\mathbb{P}\left(Z_n \leq p_{\mathrm{B}} 2^{-n} \right) \geq C(W) - c_2 2^{-n(\rho-\alpha)} = C(W) - c_2 2^{-n/\mu}, \tag{2.24}$$

where $c_2 = \sqrt{2p_{\mathrm{B}}} + 2p_{\mathrm{B}}^{-\alpha}/\delta$ and the last equality uses the definitions (2.23), (2.21) and (2.22).

Consider the transmission of a polar code of block length $N = 2^n$ and rate $R = C(W) - c_2 2^{-n/\mu}$ over W. Then, by combining (1.19) and (2.24), we have that the error probability under successive cancellation decoding is upper bounded by p_{B}. Therefore, the result (2.4) follows with $\beta_1 = c_2^{\mu}$.

A similar proof holds for the specific case in which W is a BEC.

□

Now, let us briefly sketch how to prove the result stated in Remark 2.2. First, we need to generalize Lemma 2.1 by showing that, under the same hypothesis (2.13), we have that, for any $\nu > 0$,

$$\mathbb{P}\left(Z_n \leq 2^{-n(\nu+1)} \right) \geq C(W) - c_4 2^{-n(\rho-(\nu+1)\alpha)}, \tag{2.25}$$

where $c_4 = \sqrt{2} + 2c_1$. Then, we simply follow the procedure described in the proof of Theorem 2.1 with the difference that α is a factor $1 + \nu$ smaller than in (2.21).

2.3.3 Valid Choice for Scaling Exponent

Let W be a BMS channel. The proof of Theorem 2.2 consists in providing a good candidate for the function $h(x) : [0, 1] \to [0, 1]$ such that $h(0) = h(1) = 0$, $h(x) > 0$ for any $x \in (0, 1)$ and (2.3) is satisfied with a value of μ as small as possible. In particular, we will prove that $\mu = 4.714$ is a valid choice.

The idea is to apply repeatedly the operator T_{BMSC} defined in (2.9) until we converge to the function $h(x)$. Hence, let us define $h_k(x)$ recursively for any $k \geq 1$

as

$$h_k(x) = \frac{f_k(x)}{\sup_{y\in(0,1)} f_k(y)}, \tag{2.26}$$

$$f_k(x) = \sup_{y\in[x\sqrt{2-x^2},2x-x^2]} \frac{h_{k-1}(x^2)+h_{k-1}(y)}{2}, \tag{2.27}$$

with some initial condition $h_0(x)$ such that $h_0(0) = h_0(1) = 0$ and $h_0(x) > 0$ for any $x \in (0,1)$. Note that the normalization step (2.26) ensures that the function $h_k(x)$ does not tend to the constant function 0 in the interval $[0,1]$.

However, even if we choose some simple initial condition $h_0(x)$, the sequence of functions $\{h_k(x)\}_{k\in\mathbb{N}}$ is analytically intractable. Hence, we need to resort to numerical methods, keeping in mind that we require a *provable* upper bound for any $x \in (0,1)$ on the function

$$r(x) = \sup_{y\in[x\sqrt{2-x^2},2x-x^2]} \frac{h(x^2)+h(y)}{2h(x)}. \tag{2.28}$$

To do so, first we construct an adequate candidate for the function $h(x)$. This function will depend on some auxiliary parameters. Then, we describe an algorithm to analyze this candidate and present a choice of the parameters that gives $\mu = 4.714$.

Let us underline that, although the procedure is numerical, the resulting upper bound and the value of μ are rigorously *provable*. Indeed, as we will see at the end of this section, the algorithm requires to compute the maximum of rational powers of rational numbers and this operation can be performed with arbitrary precision.

For the construction part, we observe numerically that, when k is sufficiently large, the function $h_k(x)$ depends weakly on the initial condition $h_0(x)$, and it does not change much after one more iteration, i.e., $h_{k+1}(x) \approx h_k(x)$. In addition, let us point out that the goal is *not* to obtain an exact approximation of the sequence of functions $\{h_k(x)\}_{k\in\mathbb{N}}$ defined in (2.26)-(2.27). The actual goal is to obtain a candidate $h(x)$ that satisfies (2.3) with a value of μ as low as possible.

Pick a large integer N_s and let us define the sequence of functions $\{\hat{h}_k(x)\}_{k\in\mathbb{N}}$ as follows. For any $k \in \mathbb{N}$, $\hat{h}_k(x)$ is the piece-wise linear function obtained by linear interpolation from the samples $\hat{h}_k(x_i)$, where $x_i = i/N_\mathrm{s}$ for $i \in \{0,1,\cdots,N_\mathrm{s}\}$. The samples $\hat{h}_k(x_i)$ are given by

$$\begin{aligned} \hat{h}_k(x_i) &= \frac{\hat{f}_k(x_i)}{\max_{j\in\{0,1,\cdots,N_\mathrm{s}\}} \hat{f}_k(x_j)}, \\ \hat{f}_k(x_i) &= \frac{\hat{h}_{k-1}((x_i)^2) + \max_{j\in\{0,1,\cdots,M_\mathrm{s}\}} \hat{h}_{k-1}(y_{i,j})}{2}, \end{aligned} \tag{2.29}$$

where M_s is a large integer, and, for $j \in \{0,1,\cdots,M_\mathrm{s}\}$, $y_{i,j}$ is defined as

$$y_{i,j} = x_i\sqrt{2-x_i^2} + \frac{j}{M_\mathrm{s}} x_i \left(2 - x_i - \sqrt{2-x_i^2}\right). \tag{2.30}$$

The initial samples $\hat{h}_0(x_i)$ are obtained by evaluating at the points $\{x_i\}_{i=0}^{N_\mathrm{s}}$ some function $h_0(x)$ such that $h_0(0) = h_0(1) = 0$ and $h_0(x) > 0$ for any $x \in (0,1)$ (see Figure 2.1 for a plot of $\hat{h}_0(x)$ and $\hat{h}_k(x)$).

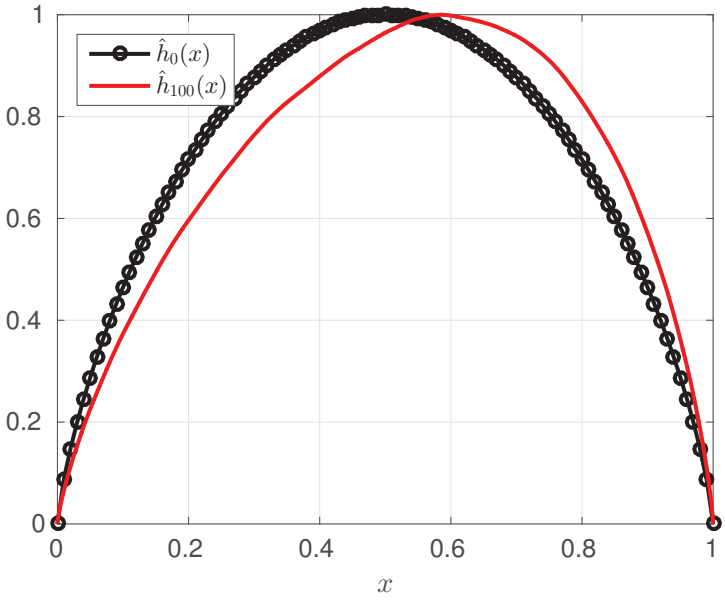


Figure 2.1 – Plot of $\hat{h}_0(x)$ (black circles) and $\hat{h}_k(x)$ (red line) after $k = 100$ steps of the recursion (2.29) with $N_\mathrm{s} = 10^6$, $M_\mathrm{s} = 10^4$, and the initial condition $f_0(x) = (x(1-x))^{3/4}$.

It is clear that, by increasing N_s and M_s, we obtain a better approximation of the sequence of functions (2.26)-(2.27). In addition, by increasing k we get closer to the limiting function $\lim_{k\to\infty} \hat{h}_k(x)$. Set

$$\hat{r}_k = \max_{i\in\{1,\cdots,N_\mathrm{s}-1\}} \frac{\hat{h}_k((x_i)^2) + \max_{j\in\{0,1,\cdots,M_\mathrm{s}\}} \hat{h}_k(y_{i,j})}{2\hat{h}_k(x_i)}. \tag{2.31}$$

We observe from numerical simulations that, when k increases, the sequence $\hat{r}_k$ tends to the limiting value 0.86275 for any k. Furthermore, this limit depends very weakly on the particular choice of the initial conditions $\{\hat{h}_0(x_i)\}_{i=0}^{N_\mathrm{s}}$.

Note that, by using the samples $\{\hat{h}_k(x_i)\}_{i=0}^{N_\mathrm{s}}$, $\hat{r}_k$ gives an indication of the smallest value of μ that we could hope for, i.e., $\mu = -1/\log_2 0.86275 = 4.695$. Indeed, if we obtain $h(x)$ by interpolating the samples $\{\hat{h}_k(x_i)\}_{i=0}^{N_\mathrm{s}}$, then $\hat{r}_k = \max_{i\in\{1,\cdots,N_\mathrm{s}-1\}} r(i/N_\mathrm{s})$, where $r(x)$ is defined in (2.28). Hence, $\hat{r}_k \le \sup_{x\in(0,1)} r(x)$, i.e., $\hat{r}_k$ is a lower bound on the desired supremum, whereas we are looking for an upper bound to that quantity.

Fix a large integer $\bar{k}$ and, before computing a provable upper bound on the quantity $\sup_{x\in(0,1)} r(x)$, let us describe the interpolation method for obtaining the candidate $h(x)$ from the samples $\{\hat{h}_{\bar{k}}(x_i)\}_{i=0}^{N_\mathrm{s}}$.

For x close to 0 and for x close to 1, linear interpolation does not yield a good candidate $h(x)$. Indeed, assume that $h(x) = \hat{h}_{\bar{k}}(x)$ for $x \in [0, 1/N_\mathrm{s}]$. Then, $\lim_{x\to 0^+} r(x) = 1$, hence $\sup_{x\in(0,1)} r(x) \ge 1$. Similarly, if $h(x) = \hat{h}_{\bar{k}}(x)$ for $x \in [1 - 1/N_\mathrm{s}, 1]$, then $\lim_{x\to 1^-} r(x) = 1$. On the contrary, if $h(x)$ grows as x^η in a neighborhood of 0 for $\eta \in (0, 1)$, then, it is easy to see that $\lim_{x\to 0^+} r(x) = 2^{\eta-1}$. Similarly, if $h(x)$ grows as $(1-x)^\eta$ in a neighborhood of 1 for $\eta \in (0, 1)$, then

$\lim_{x\to 1^-} r(x) = 2^{\eta-1}$. Consequently, the idea is to choose η slightly smaller than $1 - 1/4.695$, where 4.695 constitutes a good approximation to the target value of μ that we want to achieve. Based on this observation, we set

$$b_0(x) = \hat{h}_{\bar{k}} \left(\frac{\bar{m}}{N_\mathrm{s}}\right) \left(\frac{\bar{m}}{N_\mathrm{s}}\right)^{-\eta} x^\eta, \tag{2.32}$$

$$b_1(x) = \hat{h}_{\bar{k}} \left(1 - \frac{\bar{m}}{N_\mathrm{s}}\right) \left(\frac{\bar{m}}{N_\mathrm{s}}\right)^{-\eta} (1-x)^\eta, \tag{2.33}$$

for some integer $\bar{m} \geq 2$. Then, sample $b_0(x)$ for $x \in [1/N_\mathrm{s}, \bar{m}/N_\mathrm{s}]$, sample $\hat{h}_{\bar{k}}(x)$ for $x \in [\bar{m}/N_\mathrm{s}, 1 - \bar{m}/N_\mathrm{s}]$, and sample $b_1(x)$ for $x \in [1 - \bar{m}/N_\mathrm{s}, 1 - 1/N_\mathrm{s}]$. Note that it is better to not have a uniform sampling, but to choose the number of samples according to the rule that follows. Pick some δ_s small enough. Then, for each couple of consecutive samples, the bigger one has to be at most a factor $1 + \delta_\mathrm{s}$ larger than the smaller one. Let $\{x'_i\}_{i=1}^{N'_\mathrm{s}}$ denote the set of sampling positions and $\{\hat{h}_i\}_{i=1}^{N'_\mathrm{s}}$ denote the set of samples obtained with this procedure, where N'_s is the number of such samples. Eventually, we define the candidate $h(x)$ as

$$h(x) = \begin{cases} b_0(x), & \text{for } x \in \left[0, \dfrac{1}{N_\mathrm{s}}\right], \\ \\ b_1(x) & \text{for } x \in \left[1 - \dfrac{1}{N_\mathrm{s}}, 1\right], \end{cases} \tag{2.34}$$

and, for $x \in [1/N_\mathrm{s}, 1 - 1/N_\mathrm{s}]$, $h(x)$ is obtained by linear interpolation from the samples $\{\hat{h}_i\}$.

Concerning the analysis of $h(x)$, let us remind that the goal is to find a provable upper bound on $\sup_{x\in(0,1)} r(x)$. First, consider the values of x in a neighborhood of 0. The following chain of inequalities holds for any $x \in [0, 1/N_\mathrm{s}]$,

$$\begin{aligned} r(x) &\overset{\text{(a)}}{\leq} \frac{h(x^2) + h(2x)}{2h(x)} \\ &\overset{\text{(b)}}{\leq} \frac{b_0(x^2) + b_0(2x)}{2b_0(x)} \\ &\overset{\text{(c)}}{=} \frac{x^\eta}{2} + 2^{\eta-1} \\ &\leq H_0 \triangleq \frac{(N_\mathrm{s})^{-\eta}}{2} + 2^{\eta-1}, \end{aligned} \tag{2.35}$$

where the inequality (a) uses that $h(y) \leq h(2x)$ for any $y \in [x\sqrt{2-x^2}, 2x - x^2]$, as $h(x)$ is increasing for $x \in [0, 2/N_\mathrm{s}]$; the inequality (b) uses that $h(x) = b_0(x)$ for $x \in [0, 1/N_\mathrm{s}]$ and $h(x) \leq b_0(x)$ for $x \in [1/N_\mathrm{s}, 2/N_\mathrm{s}]$, as, in that interval, $h(x)$ is the linear interpolation of samples taken from $b_0(x)$ and $b_0(x)$ is concave for any $\eta \in (0, 1)$; and the equality (c) uses the definition (2.32) of $b_0(x)$.

Second, consider the values of x is a neighborhood of 1. The following chain of

inequalities holds for any $x \in [1 - 1/N_\mathrm{s}, 1]$,

$$\begin{aligned} r(x) &\overset{\text{(a)}}{\leq} \frac{h(x^2) + h(x\sqrt{2-x^2})}{2h(x)} \\ &\overset{\text{(b)}}{\leq} \frac{b_1(x^2) + b_1(x\sqrt{2-x^2})}{2b_1(x)} \\ &\overset{\text{(c)}}{=} \frac{(1+x)^\eta}{2} + \frac{1}{2}\left(\frac{1-x\sqrt{2-x^2}}{1-x}\right)^\eta \\ &\overset{\text{(d)}}{\leq} H_1 \triangleq 2^{\eta-1} + \frac{1}{2}\left(N_\mathrm{s} - (N_\mathrm{s}-1)\sqrt{1 + \frac{2}{N_\mathrm{s}} - \frac{1}{(N_\mathrm{s})^2}}\right)^\eta, \end{aligned} \tag{2.36}$$

where the inequality (a) uses that $h(y) \leq h(x\sqrt{2-x^2})$ for any $y \in [x\sqrt{2-x^2}, 2x - x^2]$, as $h(x)$ is decreasing for $x \in [1 - 1/N_\mathrm{s}, 1]$; the inequality (b) uses that $h(x) = b_1(x)$ for $x \in [1 - 1/N_\mathrm{s}, 1]$ and $h(x) \leq b_1(x)$ for $x \in [1/N_\mathrm{s}, 2/N_\mathrm{s}]$, as, in that interval, $h(x)$ is the linear interpolation of samples taken from $b_1(x)$ and $b_1(x)$ is concave for any $\eta \in (0,1)$; the equality (c) uses the definition (2.33) of $b_1(x)$; and the inequality (d) uses that $(1 - x\sqrt{2-x^2})(1-x)^{-1}$ is decreasing for any $x \in (0,1)$.

Finally, consider the values of x in the interval $[1/N_\mathrm{s}, 1 - 1/N_\mathrm{s}]$. For any $i \in \{1, \cdots, N'_\mathrm{s} - 1\}$, define

$$\begin{aligned} J_i^+ &= \{j : x'_j \in [(x'_i)^2, (x'_{i+1})^2]\}, \\ J_i^- &= \{j : x'_j \in [x'_i\sqrt{2-(x'_i)^2}, 2x'_{i+1} - (x'_{i+1})^2]\}. \end{aligned}$$

Then, as $h(x)$ is piece-wise linear in the interval $[1/N_\mathrm{s}, 1 - 1/N_\mathrm{s}]$, we have that, for any $x \in [x'_i, x'_{i+1}]$,

$$\begin{aligned} h(x) &\geq \min\left(h(x'_i), h(x'_{i+1})\right), \\ h(x^2) &\leq h_i^+ \triangleq \max\left(h\left((x'_i)^2\right), h\left((x'_{i+1})^2\right), \max_{j \in J_i^+}\left(h(x'_j)\right)\right), \\ \sup_{y \in [x\sqrt{2-x^2}, 2x-x^2]} h(y) &\leq h_i^- \triangleq \max\Bigg(h\left(x'_i\sqrt{2-(x'_i)^2}\right), \\ &\qquad h\left(2x'_{i+1} - (x'_{i+1})^2\right), \max_{j \in J_i^-}\left(h(x'_j)\right)\Bigg), \end{aligned}$$

which implies that, for any $x \in [x'_i, x'_{i+1}]$,

$$r(x) \leq \frac{h_i^+ + h_i^-}{2\min\left(h(x'_i), h(x'_{i+1})\right)}. \tag{2.37}$$

As a result, by combining (2.35), (2.36), and (2.37), we conclude that

$$\sup_{x \in (0,1)} r(x) \leq \max\left(H_0, H_1, \max_{i \in \{1, \cdots, N'_\mathrm{s}-1\}} \frac{h_i^+ + h_i^-}{2\min\left(h(x'_i), h(x'_{i+1})\right)}\right), \tag{2.38}$$

which implies that (2.3) holds for any μ such that $2^{-1/\mu}$ is an upper bound on the RHS of (2.38).

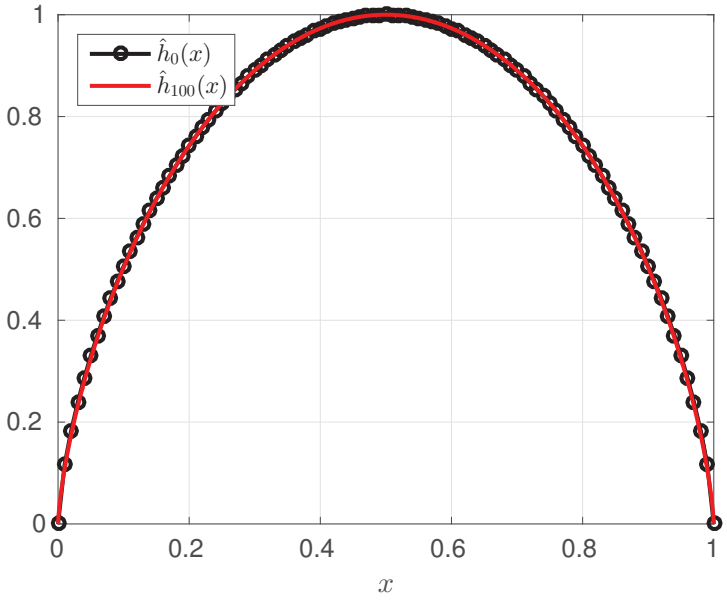


Figure 2.2 – Plot of $\hat{h}_0(x)$ (black circles) and $\hat{h}_k(x)$ (red line) after $k = 100$ steps of the recursion obtained by applying the operator T_{BEC} defined in (2.6) with $N_{\mathrm{s}} = 10^6$, $M_{\mathrm{s}} = 10^4$, and the initial condition $f_0(x) = (x(1-x))^{2/3}$. In this case, unlike in Figure 2.1, $\hat{h}_{100}(x)$ remains symmetric and very similar to the initial condition $\hat{h}_0(x)$.

Let us choose δ_{s}, η, the sampling positions $\{x'_i\}_{i=1}^{N'_{\mathrm{s}}}$, and the samples $\{\hat{h}_i\}_{i=1}^{N'_{\mathrm{s}}}$ to be rational numbers. Then, the RHS of (2.38) is the maximum of either rational numbers or sums of rational powers of rational numbers. Consequently, we can provide a provable upper bound on the RHS of (2.38), hence on μ. In particular, by setting $N_{\mathrm{s}} = 10^6$, $M_{\mathrm{s}} = 10^4$, $f_0(x) = (x(1-x))^{3/4}$, $k = 100$, $\delta_{\mathrm{s}} = 10^{-4}$, $\eta = 78/100$, and $\bar{m} = 13$, we obtain $\mu = 4.714$.

For the BEC the idea is to apply repeatedly the operator T_{BEC} defined in (2.6). Hence, by adapting the procedure described above and by setting $N_{\mathrm{s}} = 10^6$, $M_{\mathrm{s}} = 10^4$, $f_0(x) = (x(1-x))^{2/3}$, $k = 100$, $\delta_{\mathrm{s}} = 10^{-4}$, $\eta = 72/100$, and $\bar{m} = 5$, we obtain $\mu = 3.639$ (see Figure 2.2 for a plot of $\hat{h}_0(x)$ and $\hat{h}_k(x)$).

2.4 Moderate Deviations

The scaling exponent describes how fast the gap to capacity, as a function of the block length, tends to 0, when the error probability is fixed. Hence, it is natural to ask how fast the gap to capacity, as a function of the block length, tends to 0, when the error probability tends at a certain speed to 0. The discussion of Remark 2.2 in Section 2.3.1 points out that we can let the error probability go to 0 polynomially fast in N, and maintain the same scaling exponent. In this section, we show that, if we allow a less favorable scaling between gap to capacity and block length (i.e., a larger scaling exponent), then the error probability goes to 0 sub-exponentially fast in N.

2.4.1 Statement and Discussion

Theorem 2.3 (Joint Scaling: Exponential Decay of P_{B})**.** *Assume that there exists a function $h(x)$ that satisfies the hypotheses of Theorem 2.1 for some $\mu > 2$. Consider the transmission over a BMS channel W with capacity $C(W)$ by using a polar code of rate $R < C(W)$. Then, for any $\gamma \in (1/(1+\mu), 1)$, the block length N and the block error probability under successive cancellation decoding P_{B} are such that*

$$\begin{aligned} P_{\mathrm{B}} &\leq N \cdot 2^{-N^{\gamma \cdot h_2^{(-1)}\left(\frac{\gamma(\mu+1)-1}{\gamma\mu}\right)}}, \\ N &\leq \frac{\beta_3}{(C(W)-R)^{\mu/(1-\gamma)}}, \end{aligned} \tag{2.39}$$

where β_3 is a universal constant that does not depend on W or on γ, and $h_2^{(-1)}$ is the inverse of the binary entropy function defined as $h_2(x) = -x\log_2 x - (1-x)\log_2(1-x)$ for any $x \in [0, 1/2]$. If W is a BEC, the less stringent hypothesis (2.5) *on μ is required for* (2.39) *to hold.*

In short, formula (2.39) describes a trade-off between gap to capacity and error probability as functions of the block length N. Recall from Remark 2.2 that, if the scaling exponent is the μ given by Theorem 2.2, then the error probability decays polynomially fast in $1/N$. Theorem 2.3 goes one step further and proves that, in order to have a faster decay of the error probability, e.g., a sub-exponential decay, it suffices to take a larger scaling exponent.

More specifically, let γ go from $1/(1+\mu)$ to 1. On the one hand, the error probability goes faster and faster to 0, since the exponent $\gamma \cdot h_2^{(-1)}\left((\gamma(\mu+1)-1)/(\gamma\mu)\right)$ is increasing in γ; on the other hand, the gap to capacity goes slower to 0, since the exponent $\mu/(1-\gamma)$ is increasing in γ.

Before proceeding with the proof, it is useful to discuss three points. The first remark concerns the possible choices for μ in (2.39). The second remark shows how to recover from Theorem 2.3 the result [56] concerning the error exponent regime. The third remark adds the Bhattacharyya parameter $Z(W)$ to the picture outlined in Theorem 2.3 and, in particular, it focuses on the dependency between P_{B} and $Z(W)$.

Remark 2.3 (Valid Choice for μ in (2.39))**.** *By constructing a function $h(x)$ as in the proof of Theorem 2.2 contained in Section 2.3.3, we immediately have that valid choices of μ in* (2.39) *are $\mu = 4.714$ for any BMS channel and $\mu = 3.637$ for the special case of the BEC.*

Remark 2.4 (Error Exponent Regime and Theorem 2.3)**.** *By picking γ close to 1, we recover the result [56] concerning the error exponent regime: if we allow the gap to capacity to be arbitrary small but independent of N, then P_{B} is $O(2^{-N^{\beta}})$ for any $\beta \in (0, 1/2)$. Furthermore, Theorem 2.3 contains as a particular case also the stronger result in [127], where the authors prove that the block length scales polynomially fast with the inverse of the gap to capacity and the error probability can be upper bounded by $2^{-N^{0.49}}$. On the contrary, note that it is not possible to recover from Theorem 2.3 the result of Theorem 2.1 concerning the scaling exponent regime. Indeed, choose γ close to $1/(1+\mu)$. Then, the exponent $\gamma \cdot h_2^{(-1)}\left((\gamma(\mu+1)-1)/(\gamma\mu)\right)$*

tends to 0. *This means that we approach a regime in which the error probability is independent of* N, *but* N *is* $O\left(1/(C(W)-R)^{\mu+1}\right)$, *instead of* $O\left(1/(C(W)-R)^{\mu}\right)$ *as in* (2.4). *We believe that this is only an artifact of the proof technique used to show Theorem 2.3 and that it might be possible to find a joint scaling that contains as special cases the error exponent and the scaling exponent regimes.*

Remark 2.5 (Dependency between P_{B} and $Z(W)$)**.** *Consider the transmission over a BMS channel* W *with Bhattacharyya parameter* $Z(W)$. *Then, under the hypotheses of Theorem 2.3, it is possible to prove that*

$$\begin{aligned} P_{\mathrm{B}} &\leq N \cdot Z(W)^{\frac{1}{2}\cdot N^{\gamma \cdot h_2^{(-1)}\left(\frac{\gamma(\mu+1)-1}{\gamma\mu}\right)}}, \\ N &\leq \frac{\beta_4}{(C(W)-R)^{\mu/(1-\gamma)}}, \end{aligned} \tag{2.40}$$

where β_4 *is a universal constant that does not depend on* W *or on* γ. *A sketch of the proof of this statement is given in Appendix 2.6.3. This result means that the error probability scales as* $Z(W)$ *raised to some power of* N, *where the exponent follows the trade-off of Theorem 2.3. To see that this bound is meaningful, consider the case of the transmission over the BEC in the error exponent regime. On the one hand, formula* (2.40) *gives that* P_{B} *scales roughly as* $Z(W)^{\sqrt{N}}$. *On the other hand,* $P_{\mathrm{B}} \geq \max_{i\in\mathcal{I}} Z_n^{(i)}$, *where* $\mathcal{I}$ *denotes the set of information positions and* $Z_n^{(i)}$ *is a polynomial in* $Z(W)$ *with minimum degree that scales roughly*[2] *as* $\sqrt{N}$. *The scaling between the error probability and the Bhattacharyya parameter will be further explored in Section 2.5.*

2.4.2 Proof of Theorem 2.3

Proof. Let $Z_n(W)$ be the Bhattacharyya process associated with the channel W and defined in (1.11). Then, by following the same procedure that gives (2.24), we have that, for any $n_0 \in \mathbb{N}$,

$$\mathbb{P}\left(Z_{n_0} \leq 2^{-n_0}\right) \geq C(W) - c_5\, 2^{-n_0/\mu}, \tag{2.41}$$

where c_5 is a constant that does not depend on n and is given by $c_5 = \sqrt{2} + 2/\delta$, with δ defined in (2.22).

Let $\{B_n\}_{n\geq 1}$ be a sequence of i.i.d. Bernoulli(1/2) random variables. Then, by using (1.11), it is clear that, for $n \geq 1$,

$$Z_{n_0+n} \leq \begin{cases} Z_{n_0+n-1}^2, & \text{if } B_n = 1, \\ 2Z_{n_0+n-1}, & \text{if } B_n = 0. \end{cases}$$

Therefore, by applying Lemma 22 of [128], we obtain that, for $n_1 \geq 1$,

$$\mathbb{P}\left(Z_{n_0+n_1} \leq 2^{-2^{\sum_{i=1}^{n_1} B_i}} \mid Z_{n_0} = x\right) \geq 1 - c_6\, x(1-\log_2 x), \tag{2.42}$$

[2]Note that the minimum degree of $Z_n^{(i)}$ seen as a polynomial in $Z(W)$ is equal to the minimum distance of the code and that the minimum distance scales roughly as $\sqrt{N}$ according to Lemma 4 of [53].

with $c_6 = 2/(\sqrt{2}-1)^2$.

Consequently, we have that

$$\begin{aligned}\mathbb{P}\left(Z_{n_0+n_1} \le 2^{-2^{\sum_{i=1}^{n_1} B_i}}\right) &= \mathbb{P}\left(Z_{n_0} \le 2^{-n_0}\right) \\ &\quad \cdot \mathbb{P}\left(Z_{n_0+n_1} \le 2^{-2^{\sum_{i=1}^{n_1} B_i}} \mid Z_{n_0} \le 2^{-n_0}\right) \\ &\overset{\text{(a)}}{\ge} \mathbb{P}\left(Z_{n_0} \le 2^{-n_0}\right) \cdot \left(1 - c_6 2^{-n_0}(1+n_0)\right) \\ &\overset{\text{(b)}}{\ge} \left(C(W) - c_5 2^{-n_0/\mu}\right) \cdot \left(1 - c_6 \frac{\sqrt{2}}{\ln 2} 2^{-n_0/2}\right) \\ &\overset{\text{(c)}}{\ge} C(W) - \left(c_5 + c_6 \frac{\sqrt{2}}{\ln 2}\right) 2^{-n_0/\mu},\end{aligned} \tag{2.43}$$

where the inequality (a) uses (2.42) and the fact that $1 - c_6 x(1-\log_2 x)$ is decreasing in x for any $x \le 2^{-n_0} \le 1/2$; the inequality (b) uses (2.41) and that $1 - c_6 2^{-n_0}(1+n_0) \ge 1 - c_6 2^{-n_0/2}\sqrt{2}/\ln 2$ for any $n_0 \in \mathbb{N}$; and the inequality (c) uses that $\mu > 2$.

Let $h_2(x) = -x\log_2 x - (1-x)\log_2(1-x)$ denote the binary entropy function. Then, for any $\epsilon \in (0, 1/2)$,

$$\begin{aligned}\mathbb{P}\left(2^{-2^{\sum_{i=1}^{n_1} B_i}} > 2^{-2^{n_1\epsilon}}\right) &= \mathbb{P}\left(\sum_{i=1}^{n_1} B_i < n_1\epsilon\right) \\ &\le \mathbb{P}\left(\sum_{i=1}^{n_1} B_i \le \lfloor n_1\epsilon \rfloor\right) \\ &= \sum_{k=0}^{\lfloor n_1\epsilon\rfloor} \binom{n_1}{k} \left(\frac{1}{2}\right)^{n_1} \\ &\overset{\text{(a)}}{\le} \left(\frac{1}{2}\right)^{n_1} 2^{n_1 h_2(\lfloor n_1\epsilon\rfloor/n_1)} \\ &\overset{\text{(b)}}{\le} 2^{-n_1(1-h_2(\epsilon))},\end{aligned} \tag{2.44}$$

where the inequality (a) uses formula (1.59) of [44]; and the inequality (b) uses that $h_2(x)$ is increasing for any $x \le 1/2$.

Note that, for any two events A and B, $\mathbb{P}(A \cap B) \ge \mathbb{P}(A) + \mathbb{P}(B) - 1$. Hence, by combining (2.43) and (2.44), we obtain that

$$\mathbb{P}\left(Z_{n_0+n_1} \le 2^{-2^{n_1\epsilon}}\right) \ge C(W) - \left(c_5 + c_6 \frac{\sqrt{2}}{\ln 2}\right) 2^{-n_0/\mu} - 2^{-n_1(1-h_2(\epsilon))}. \tag{2.45}$$

Let $n \ge 1$. Set $n_1 = \lceil \gamma n \rceil$, $n_0 = n - \lceil \gamma n \rceil$, and $\epsilon = h_2^{(-1)}\left((\gamma(\mu+1)-1)/(\gamma\mu)\right)$, where $h_2^{(-1)}(\cdot)$ is the inverse of $h_2(x)$ for any $x \in [0, 1/2]$. Note that if $\gamma \in (1/(1+\mu), 1)$, then $\epsilon \in (0, 1/2)$. Consequently, formula (2.45) can be rewritten as

$$\mathbb{P}\left(Z_{n_0+n_1} \le 2^{-2^{n\gamma h_2^{(-1)}\left(\frac{\gamma(\mu+1)-1}{\gamma\mu}\right)}}\right) \ge C(W) - c_7 2^{-n\frac{1-\gamma}{\mu}}, \tag{2.46}$$

with $c_7 = 1 + \sqrt{2}\left(c_5 + c_6\sqrt{2}/\ln 2\right)$.

Consider the transmission of a polar code of block length $N = 2^n$ and rate R given by the RHS of (2.46). Then, the result (2.39) holds with $\beta_3 = c_7^{\mu}$.

□

2.5 Absence of Error Floors

In the discussion of Remark 2.5 in Section 2.4.1, we study the dependency between the error probability and the Bhattacharyya parameter, and we consider a setting in which, as the channel varies, the polar code used for the transmission changes accordingly. In this section, we consider a different scenario in which the polar code stays fixed as the channel varies, and we prove a result about the speed of decay of the error probability as a function of the Bhattacharyya parameter of the channel. By doing so, we conclude that polar codes are not affected by error floors.

2.5.1 Statement and Discussion

Let $\mathcal{C}$ be the polar code with information set $\mathcal{I}$ designed for the transmission over the BMS channel W' with Bhattacharyya parameter $Z(W')$. Then, the actual channel over which the transmission takes place is the BMS channel W with Bhattacharyya parameter $Z(W)$. In the error floor regime, the code $\mathcal{C}$ is fixed and W varies. The aim is to study the scaling between the error probability under SC decoding and the Bhattacharyya parameter $Z(W)$.

Denote by $Z_n^{(i)}(W)$ the Bhattacharyya parameter of the synthetic channel of index i obtained from W after n steps of polarization. The main result is presented in Theorem 2.4 and it relates $Z_n^{(i)}(W)$ obtained from W to $Z_n^{(i)}(W')$ obtained from W'. From this, in Corollary 2.1, we relate the sum of the Bhattacharyya parameters at the information positions obtained from W, i.e., $\tilde{P}_{\mathrm{B}}(W) \triangleq \sum_{i\in\mathcal{I}} Z_n^{(i)}(W)$, to the sum of Bhattacharyya parameters obtained from W', i.e., $\tilde{P}_{\mathrm{B}}(W') \triangleq \sum_{i\in\mathcal{I}} Z_n^{(i)}(W')$. Note that the indices of the information positions are the same in both sums, since the information set $\mathcal{I}$ is fixed. The proof of Theorem 2.4 is in Section 2.5.2, and the proof of Corollary 2.1 naturally follows.

Theorem 2.4 (Scaling of $Z_n^{(i)}(W)$)**.** *Consider two BMS channels W and W' with Bhattacharyya parameter $Z(W)$ and $Z(W')$, respectively. For $n \in \mathbb{N}$ and $i \in \{1, \cdots, 2^n\}$, let $Z_n^{(i)}(W)$ be the Bhattacharyya parameter of the channel $W_n^{(i)}$ obtained from W via channel polarization and let $Z_n^{(i)}(W')$ be similarly obtained from W'. If $Z(W) \leq Z(W')^2$, then*

$$Z_n^{(i)}(W) \leq Z_n^{(i)}(W')^{\frac{\log_2 Z(W)}{\log_2 Z(W')}}. \tag{2.47}$$

If W and W' are BECs, then (2.47) holds if $Z(W) \leq Z(W')$.

Corollary 2.1 (Scaling of $\tilde{P}_{\mathrm{B}}(W)$)**.** *Let W' be a BMS channel with Bhattacharyya parameter $Z(W')$ and let $\mathcal{C}$ be the polar code of block length $N = 2^n$ and rate R for the transmission over W'. Denote by $\tilde{P}_{\mathrm{B}}(W')$ the sum of the Bhattacharyya parameters at the information positions obtained from W', i.e., $\tilde{P}_{\mathrm{B}}(W') \triangleq \sum_{i\in\mathcal{I}} Z_n^{(i)}(W')$, where*

$\mathcal{I}$ is the information set of the polar code $\mathcal{C}$. Now, consider the transmission over the BMS channel W with Bhattacharyya parameter $Z(W)$ by using the polar code $\mathcal{C}$ and let $\tilde{P}_{\mathrm{B}}(W)$ be the sum of the Bhattacharyya parameters at the information positions obtained from W, i.e., $\tilde{P}_{\mathrm{B}}(W) \triangleq \sum_{i\in\mathcal{I}} Z_n^{(i)}(W)$. If $Z(W) \leq Z(W')^2$, then

$$\tilde{P}_{\mathrm{B}}(W) \leq \tilde{P}_{\mathrm{B}}(W')^{\frac{\log_2 Z(W)}{\log_2 Z(W')}}. \tag{2.48}$$

If W and W' are BECs, then (2.48) holds if $Z(W) \leq Z(W')$.

Now, let us discuss how the results above imply that polar codes are not affected by error floors. Denote by $P_{\mathrm{B}}(W)$ the error probability under SC decoding for transmission of $\mathcal{C}$ over W and recall from (1.19) that $P_{\mathrm{B}}(W) \leq \tilde{P}_{\mathrm{B}}(W)$. Hence, formula (2.48) implies that

$$P_{\mathrm{B}}(W) \leq Z(W)^{\frac{\log_2 \tilde{P}_{\mathrm{B}}(W')}{\log_2 Z(W')}}. \tag{2.49}$$

Note that the upper bound (2.40) on P_{B} comes from an identical upper bound on the sum of the Bhattacharyya parameters $\tilde{P}_{\mathrm{B}}$. Thus, by choosing $\gamma \approx 1$ in (2.40), we have that $\tilde{P}_{\mathrm{B}}(W')$ scales roughly as $Z(W')^{\sqrt{N}}$. Therefore, from (2.49) we conclude that $P_{\mathrm{B}}(W)$ scales roughly as $Z(W)^{\sqrt{N}}$. This fact excludes the existence of an error floor region.

Furthermore, in the discussion of Remark 2.5, we pointed out that $P_{\mathrm{B}}(W)$ scales as $Z(W)^{\sqrt{N}}$. This result holds when W is fixed and the polar code can be constructed according to the actual transmission channel. Whereas, in the error floor regime, we fix a polar code and we let the transmission channel vary, which means that the code cannot depend on the transmission channel. Hence, from the discussion above, it follows that the dependency between the error probability and the Bhattacharyya parameter of the channel is essentially the same as in the case in which we design the polar code for the actual transmission channel. As a result, in terms of this particular scaling, nothing is lost by considering a "mismatched" code. However, considering a "mismatched" code yields a loss in rate. Indeed, if W and W' are BECs, then (1.4) holds with equality and $Z(W) \leq Z(W')$ implies that $C(W) \geq C(W')$. If W and W' can be any BMS channels, by using (1.4) and (1.5) we easily deduce that $Z(W) \leq Z(W')^2$ implies $C(W) \geq C(W')$. Recall that the rate of a polar code for W' is such that $R < C(W')$, and the rate of a polar code for W is such that $R < C(W)$. As $C(W) \geq C(W')$, by constructing a polar code for W, we can transmit reliably at larger rates.

Before proceeding with the proof of Theorem 2.4, let us make a brief remark concerning the case $Z(W) \in \big(Z(W')^2, Z(W')\big]$.

Remark 2.6 (The Case $Z(W) \in (Z(W')^2, Z(W')]$)**.** *If W and W' are BECs, then (2.47) and (2.48) hold for any $Z(W) \leq Z(W')$, i.e., for the whole range of parameters of interest, as we think of W as a "better" channel than W'. On the contrary, if W and W' can be any BMS channels, we require that $Z(W) \leq Z(W')^2$. If there is no additional hypothesis on W and W', the main result (2.47) cannot hold in the case $Z(W) \in (Z(W')^2, Z(W')]$. Indeed, if $Z(W) = Z(W')$, we can choose W and W' such that $C(W) < C(W')$. If $C(W) < C(W')$, then the number of indices i_1 such that $\lim_{n\to\infty} Z_n^{(i_1)}(W) = 0$ is smaller than the number of*

indices i_2 such that $\lim_{n\to\infty} Z_n^{(i_2)}(W') = 0$. Hence, (2.47) cannot hold for any $i \in \{1, \cdots, 2^n\}$. A natural additional hypothesis consists in assuming that W' is degraded with respect to W, i.e., $W \succ W'$. In this case, we can at least ensure that $Z_n^{(i)}(W) \leq Z_n^{(i)}(W')$. However, it is possible to find W and W' such that (2.47) is violated for $n = 1$ when $Z(W) \in (Z(W')^2, Z(W')]$. We leave as open questions whether the bound (2.48) is still valid and what kind of looser bound holds, when $W \succ W'$ and $Z(W) \in (Z(W')^2, Z(W')]$.

2.5.2 Proof of Theorem 2.4

Proof. Assume that, for any $j \in \{1, \cdots, 2^{n-1}\}$ and for some $\eta \in \mathbb{R}^+$,

$$Z_{n-1}^{(j)}(W) \leq Z_{n-1}^{(j)}(W')^{\eta}. \tag{2.50}$$

Then, let us study for what values of η we have that (2.50) implies that, for any $i \in \{1, \cdots, 2^n\}$,

$$Z_n^{(i)}(W) \leq Z_n^{(i)}(W')^{\eta}. \tag{2.51}$$

First, consider the case in which the last polarization step is a "+" step, i.e.,

$$W_n^{(i)} = \left(W_{n-1}^{(i^+)}\right)^+, \tag{2.52}$$

for some index $i^+ \in \{1, \cdots, 2^{n-1}\}$. Hence, the following chain of inequalities holds for any BMS channel W:

$$\begin{aligned} Z_n^{(i)}(W) &\overset{\text{(a)}}{=} \left(Z_{n-1}^{(i^+)}(W)\right)^2 \\ &\overset{\text{(b)}}{\leq} \left(Z_{n-1}^{(i^+)}(W')\right)^{2\eta} \\ &\overset{\text{(c)}}{=} \left(Z_n^{(i)}(W')\right)^{\eta}, \end{aligned} \tag{2.53}$$

where the equality (a) uses (1.8); the inequality (b) uses the assumption (2.50) with $j = i^+$; and the equality (c) uses again (1.8). Consequently, if the last polarization step is a "+" step, then (2.51) holds for any BMS channel W without any restriction on η.

Then, consider the case in which the last polarization step is a "−" step, i.e.,

$$W_n^{(i)} = \left(W_{n-1}^{(i^-)}\right)^-, \tag{2.54}$$

for some index $i^- \in \{1, \cdots, 2^{n-1}\}$. Hence, the following chain of inequalities holds for any BMS channel W:

$$\begin{aligned} Z_n^{(i)}(W) &\overset{\text{(a)}}{\leq} Z_{n-1}^{(i^-)}(W)\left(2 - Z_{n-1}^{(i^-)}(W)\right) \\ &\overset{\text{(b)}}{\leq} \left(Z_{n-1}^{(i^-)}(W')\right)^{\eta}\left(2 - \left(Z_{n-1}^{(i^-)}(W')\right)^{\eta}\right) \\ &\overset{\text{(c)}}{\leq} \left(Z_{n-1}^{(i^-)}(W')\right)^{\eta}\left(2 - \left(Z_{n-1}^{(i^-)}(W')\right)^{2}\right)^{\eta/2} \\ &\overset{\text{(d)}}{\leq} \left(Z_n^{(i)}(W')\right)^{\eta}, \end{aligned} \tag{2.55}$$

where the inequality (a) uses (1.7); the inequality (b) uses the assumption (2.50) with $j = i^-$; the inequality (c) uses that $2 - x^\eta \leq (2 - x^2)^{\eta/2}$ for any $x \in [0, 1]$ if and only if $\eta \geq 2$; and the inequality (d) uses again (1.7). Consequently, if the last polarization step is a "$-$" step, then (2.51) holds for any BMS channel W, provided that $\eta \geq 2$. If W is a BEC, a less restrictive condition on η is necessary. Indeed, the following chain of inequalities holds when W is a BEC:

$$\begin{aligned} Z_n^{(i)}(W) &\overset{\text{(a)}}{=} Z_{n-1}^{(i^-)}(W)\left(2 - Z_{n-1}^{(i^-)}(W)\right) \\ &\overset{\text{(b)}}{\leq} \left(Z_{n-1}^{(i^-)}(W')\right)^\eta \left(2 - \left(Z_{n-1}^{(i^-)}(W')\right)^\eta\right) \\ &\overset{\text{(c)}}{\leq} \left(Z_{n-1}^{(i^-)}(W')\right)^\eta \left(2 - Z_{n-1}^{(i^-)}(W')\right)^\eta \\ &\overset{\text{(d)}}{=} \left(Z_n^{(i)}(W')\right)^\eta, \end{aligned} \tag{2.56}$$

where the equality (a) uses (1.9); the inequality (b) uses the assumption (2.50) with $j = i^-$; the inequality (c) uses that $2 - x^\eta \leq (2 - x)^\eta$ for any $x \in [0, 1]$ if and only if $\eta \geq 1$; and the equality (d) uses again (1.9). Consequently, if the last polarization step is a "$-$" step and W is a BEC, then (2.51) holds provided that $\eta \geq 1$.

By combining (2.53) and (2.55), we have that if (2.50) holds for $\eta \geq 2$ after $n - 1$ steps of polarization, then the same relation holds for $\eta \geq 2$ after n steps of polarization. This means that the inequality stays preserved after one more step of polarization. Clearly, as the Bhattacharyya parameter is between 0 and 1, a smaller value of η gives a tighter bound. As $Z_0^{(1)}(W) = Z(W)$ and $Z_0^{(1)}(W') = Z(W')$, the smallest choice for η is $\log_2 Z(W)/\log_2 Z(W')$. The condition $\eta \geq 2$ is equivalent to $Z(W) \leq Z(W')^2$ and, for the case of the BEC, the condition $\eta \geq 1$ is equivalent to $Z(W) \leq Z(W')$. Eventually, the result (2.47) follows easily by induction.

□

2.6 Appendix

2.6.1 Proof of Lemma 2.1

Proof. First of all, we upper bound $\mathbb{P}(Z_n \in [p_{\mathrm{B}} 2^{-n}, 1 - p_{\mathrm{B}} 2^{-n}])$ as follows:

$$\begin{aligned} \mathbb{P}\left(Z_n \in \left[p_{\mathrm{B}} 2^{-n}, 1 - p_{\mathrm{B}} 2^{-n}\right]\right) &\overset{\text{(a)}}{=} \mathbb{P}\left((Z_n(1 - Z_n))^\alpha \geq (p_{\mathrm{B}} 2^{-n}(1 - p_{\mathrm{B}} 2^{-n}))^\alpha\right) \\ &\overset{\text{(b)}}{\leq} \frac{\mathbb{E}\left[(Z_n(1 - Z_n))^\alpha\right]}{(p_{\mathrm{B}} 2^{-n}(1 - p_{\mathrm{B}} 2^{-n}))^\alpha} \\ &\overset{\text{(c)}}{\leq} \frac{c_1 2^{-n\rho}}{(p_{\mathrm{B}} 2^{-n}(1 - p_{\mathrm{B}} 2^{-n}))^\alpha} \\ &\overset{\text{(d)}}{\leq} 2 c_1 p_{\mathrm{B}}^{-\alpha} 2^{-n(\rho - \alpha)}, \end{aligned} \tag{2.57}$$

where the equality (a) uses the concavity of the function $f(x) = (x(1 - x))^\alpha$; the inequality (b) follows from Markov inequality; the inequality (c) uses the hypothesis

$\mathbb{E}[(Z_n(1-Z_n))^\alpha] \leq c_1 2^{-n\rho}$; and the inequality (d) uses that $1 - p_{\mathrm{B}} 2^{-n} \geq 1/2$ for any $n \geq 1$.

Let us define

$$\begin{aligned} A &= \mathbb{P}\left(Z_n \in \left[0, p_{\mathrm{B}} 2^{-n}\right)\right), \\ B &= \mathbb{P}\left(Z_n \in \left[p_{\mathrm{B}} 2^{-n}, 1 - p_{\mathrm{B}} 2^{-n}\right]\right), \\ C &= \mathbb{P}\left(Z_n \in \left(1 - p_{\mathrm{B}} 2^{-n}, 1\right]\right), \end{aligned} \tag{2.58}$$

and let A', B', and C' be the fraction of A, B, and C, respectively, that will go to 0 as $n \to \infty$. More formally,

$$\begin{aligned} A' &= \liminf_{m\to\infty} \mathbb{P}\left(Z_n \in \left[0, p_{\mathrm{B}} 2^{-n}\right), Z_{n+m} \leq 2^{-m}\right), \\ B' &= \liminf_{m\to\infty} \mathbb{P}\left(Z_n \in \left[p_{\mathrm{B}} 2^{-n}, 1 - p_{\mathrm{B}} 2^{-n}\right], Z_{n+m} \leq 2^{-m}\right), \\ C' &= \liminf_{m\to\infty} \mathbb{P}\left(Z_n \in \left(1 - p_{\mathrm{B}} 2^{-n}, 1\right], Z_{n+m} \leq 2^{-m}\right). \end{aligned} \tag{2.59}$$

Note that we only need that Z_{n+m} goes to 0 as m goes large, and we do not have any requirement on the speed at which it does so. Hence, we could substitute 2^{-m} in (2.59) with any other function that goes to 0 as $m \to \infty$ and that is $\Theta(2^{-2^{\beta m}})$ for some $\beta \in (0, 1/2)$, see [56].

It is clear that

$$A' + B' + C' = \liminf_{m\to\infty} \mathbb{P}\left(Z_{n+m} \leq 2^{-m}\right) = C(W). \tag{2.60}$$

In addition, from (2.57), we have that

$$B' \leq B \leq 2\, c_1\, p_{\mathrm{B}}^{-\alpha}\, 2^{-n(\rho-\alpha)}. \tag{2.61}$$

In order to upper bound C', we proceed as follows:

$$\begin{aligned} C' &= \liminf_{m\to\infty} \mathbb{P}\left(Z_{n+m} \leq 2^{-m} \mid Z_n \in \left(1 - p_{\mathrm{B}} 2^{-n}, 1\right]\right) \cdot \mathbb{P}\left(Z_n \in \left(1 - p_{\mathrm{B}} 2^{-n}, 1\right]\right) \\ &\leq \liminf_{m\to\infty} \mathbb{P}\left(Z_{n+m} \leq 2^{-m} \mid Z_n \in \left(1 - p_{\mathrm{B}} 2^{-n}, 1\right]\right). \end{aligned} \tag{2.62}$$

The last term equals the capacity of a channel with Bhattacharyya parameter in the interval $(1 - p_{\mathrm{B}} 2^{-n}, 1]$. By using (1.5), we obtain that

$$C' \leq \sqrt{1 - (1 - p_{\mathrm{B}} 2^{-n})^2} \leq \sqrt{2 p_{\mathrm{B}} 2^{-n}}. \tag{2.63}$$

As a result, we have that

$$\begin{aligned} \mathbb{P}\left(Z_n \in \left[0, p_{\mathrm{B}} 2^{-n}\right)\right) = A &\geq A' \\ &\overset{\text{(a)}}{=} C(W) - B' - C' \\ &\overset{\text{(b)}}{\geq} C(W) - 2\, c_1\, p_{\mathrm{B}}^{-\alpha}\, 2^{-n(\rho-\alpha)} - \sqrt{2 p_{\mathrm{B}} 2^{-n}}, \\ &\overset{\text{(c)}}{\geq} C(W) - \left(\sqrt{2 p_{\mathrm{B}}} + 2\, c_1\, p_{\mathrm{B}}^{-\alpha}\right) 2^{-n(\rho-\alpha)}, \end{aligned}$$

where the equality (a) uses (2.60); the inequality (b) uses (2.61) and (2.63); and the inequality (c) uses that $\rho \leq 1/2$. This chain of inequalities implies the desired result.

□

2.6.2 Proof of Lemma 2.2

Proof. Let $\alpha^* = \min(1/2, \rho_1/\log_2(4/3))$. As $\mathbb{E}\left[\left(Z_n(1-Z_n)\right)^\alpha\right]$ is decreasing in α, we can assume that $\alpha < \alpha^*$ without loss of generality. As $h(x) \geq 0$ for any $x \in [0,1]$ and $Z_n \in [0,1]$ for any $n \in \mathbb{N}$, we have that

$$\mathbb{E}\left[\left(Z_n(1-Z_n)\right)^\alpha\right] \leq \frac{1}{\delta}\mathbb{E}\left[(1-\delta)h(Z_n) + \delta(Z_n(1-Z_n))^\alpha\right] = \frac{1}{\delta}\mathbb{E}\left[g(Z_n)\right], \quad (2.64)$$

with

$$g(x) = (1-\delta)h(x) + \delta(x(1-x))^\alpha. \quad (2.65)$$

Let

$$L_g = \sup_{x\in(0,1), y\in[x\sqrt{2-x^2}, 2x-x^2]} \frac{g(x^2)+g(y)}{2g(x)}.$$

Then, by definition (1.11) of the Bhattacharyya process Z_n, we have that

$$\mathbb{E}\left[g(Z_n) \mid Z_{n-1}\right] \leq g(Z_{n-1})L_g.$$

Consequently, by induction, we can readily prove that

$$\mathbb{E}\left[g(Z_n)\right] \leq (L_g)^n g(Z(W)) \leq (L_g)^n, \quad (2.66)$$

where the last inequality follows from the fact that $g(x) \leq 1$ for $x \in [0,1]$.

Now, by combining (2.64) with (2.66), we obtain that

$$\mathbb{E}[(Z_n(1-Z_n))^\alpha] \leq \frac{1}{\delta}(L_g)^n. \quad (2.67)$$

Hence, to conclude the proof it remains to find an upper bound on L_g, i.e., to show that $L_g \leq 2^{-\rho_1} + 2\sqrt{2}\delta c_3$. By using (2.15), after some calculations, we have that

$$\frac{g(x^2)+g(y)}{2g(x)} \leq \frac{(1-\delta)h(x)2^{-\rho_1} + \frac{\delta}{2}\Big(\left(x^2(1-x)(1+x)\right)^\alpha + (y(1-y))^\alpha\Big)}{(1-\delta)h(x) + \delta(x(1-x))^\alpha}. \quad (2.68)$$

For any $y \in [x\sqrt{2-x^2}, 2x-x^2]$, we obtain

$$y(1-y) \leq x(2-x)(1-x\sqrt{2-x^2}). \quad (2.69)$$

In addition, for any $x \in (0,1)$,

$$1 - x\sqrt{2-x^2} \leq (1-x)^{4/3}. \quad (2.70)$$

In order to prove (2.70), one strategy is the following: elevate the LHS and the RHS to the third power; isolate on one side the terms that multiply $\sqrt{2-x^2}$; and square again the LHS and the RHS. In this way, we have that (2.70) is equivalent to

$$(1-x)^4(2+8x+3x^2+4x^3-4x^4-4x^5-x^6) \geq 0,$$

which is clearly satisfied when $x \in (0,1)$.

Therefore, by combining (2.68), (2.69), and (2.70), we obtain that

$$\frac{g(x^2)+g(y)}{2g(x)} \leq \frac{(1-\delta)h(x)2^{-\rho_1} + \delta(x(1-x))^\alpha t(x)}{(1-\delta)h(x) + \delta(x(1-x))^\alpha}, \quad (2.71)$$

with

$$t(x) = \frac{1}{2}\left(\big(x(1+x)\big)^{\alpha} + \big((2-x)(1-x)^{1/3}\big)^{\alpha}\right). \tag{2.72}$$

First of all, we upper bound the expression on the RHS of (2.71) when x is small. Clearly, $t(0) < 2^{-\rho_1}$ and $t(1/2) > 2^{-\rho_1}$, as $\rho_1 \leq 1/2$ and $\alpha < \alpha^*$. In addition, some passages of calculus show that the second derivative of $t(x)$ is given by

$$\begin{aligned}&\frac{\alpha}{2}\frac{(x(1+x))^{\alpha}}{x^2(1+x)^2}\left(-1-2x-2x^2+\alpha(1+2x)^2\right)\\&\quad+\frac{\alpha}{18}\frac{\big((2-x)(1-x)^{1/3}\big)^{\alpha}}{(2-3x+x^2)^2}\left(-21+30x-12x^2+\alpha(5-4x)^2\right).\end{aligned}$$

As $\alpha < 1/2$, we have that

$$\begin{aligned}-1-2x-2x^2+\alpha(1+2x)^2 &\leq -1-2x-2x^2+\frac{(1+2x)^2}{2} < 0,\\ -21+30x-12x^2+\alpha(5-4x)^2 &\leq -1-2x-2x^2+\frac{(5-4x)^2}{2} < 0.\end{aligned} \tag{2.73}$$

Hence, $t(x)$ is concave for any $x \in (0,1)$. This implies that there exist $\epsilon_1(\alpha), \epsilon_2(\alpha) \in (0,1)$ such that

$$t(x) \leq 2^{-\rho_1}, \qquad \forall\, x \in [0, \epsilon_1(\alpha)] \cup [1-\epsilon_2(\alpha), 1]. \tag{2.74}$$

Indeed, the precise values of $\epsilon_1(\alpha)$ and $\epsilon_2(\alpha)$ can be found from (2.18). By combining (2.71) with (2.74), we have that, for any $x \in [0, \epsilon_1(\alpha)] \cup [1-\epsilon_2(\alpha), 1]$ and for any $y \in [x\sqrt{2-x^2}, 2x-x^2]$,

$$\frac{g(x^2)+g(y)}{2g(x)} \leq 2^{-\rho_1}. \tag{2.75}$$

Then, we upper bound the expression on the RHS of (2.71) when x is not too small, namely, $x \in (\epsilon_1(\alpha), 1-\epsilon_2(\alpha))$:

$$\begin{aligned}\frac{(1-\delta)h(x)2^{-\rho_1}+\delta(x(1-x))^{\alpha}t(x)}{(1-\delta)h(x)+\delta(x(1-x))^{\alpha}} &\overset{\text{(a)}}{\leq} \frac{(1-\delta)h(x)2^{-\rho_1}+\delta(x(1-x))^{\alpha}2^{\alpha}}{(1-\delta)h(x)+\delta(x(1-x))^{\alpha}}\\ &\overset{\text{(b)}}{\leq} 2^{-\rho_1}+\delta\frac{2^{\alpha}}{1-\delta}\frac{(x(1-x))^{\alpha}}{h(x)}\\ &\overset{\text{(c)}}{\leq} 2^{-\rho_1}+\sqrt{2}\frac{\delta}{1-\delta}c_3,\end{aligned} \tag{2.76}$$

where the inequality (a) uses that $t(x) \leq 2^{\alpha}$ for any $x \in (0,1)$; the inequality (b) uses that $h(x) \geq 0$ and $(x(1-x))^{\alpha} \geq 0$; and the inequality (c) uses that $\alpha \leq 1/2$, and the definition of c_3 in (2.17). By putting (2.75) and (2.76) together, we have that

$$L_g \leq 2^{-\rho_1}+\sqrt{2}\frac{\delta}{1-\delta}c_3. \tag{2.77}$$

By combining (2.67) and (2.77), the result for a general BMS channel follows.

Finally, consider the special case in which W is a BEC. Clearly, (2.64) still holds, and, by using the definition (1.2) of the Bhattacharyya process Z_n for the BEC, in analogy to (2.66), we obtain that

$$\mathbb{E}[(Z_n(1-Z_n))^{\alpha}] \leq \frac{1}{\delta}(L'_g)^n, \tag{2.78}$$

where we define

$$L'_g = \sup_{x\in(0,1)} \frac{g(x^2)+g(2x-x^2)}{2g(x)}.$$

By using (2.19), after some calculations, we have that

$$\frac{g_0(x^2)+g_0(2x-x^2)}{2g_0(x)} \leq \frac{(1-\delta)h(x)2^{-\rho_1}+\delta(x(1-x))^{\alpha}t'(x)}{(1-\delta)h(x)+\delta(x(1-x))^{\alpha}},$$

with

$$t'(x) = \frac{1}{2}\left(\big(x(1+x)\big)^{\alpha} + \big((2-x)(1-x)\big)^{\alpha}\right).$$

As $(1-x) \leq (1-x)^{1/3}$ for any $x \in (0,1)$, we obtain that $t'(x) \leq t(x)$, with $t(x)$ defined in (2.72). Therefore, the result for the BEC naturally follows. □

2.6.3 Sketch of the Proof of (2.40)

Eventually, let us briefly sketch how to prove the result stated in Remark 2.5. The dependency on the Bhattacharyya parameter $Z(W)$ first appears in formula (2.66). Hence, under the hypothesis of Lemma 2.2, we can easily prove that

$$\mathbb{E}\left[(Z_n(1-Z_n))^{\alpha}\right] \leq \frac{g(Z(W))}{\delta}\left(2^{-\rho_1}+\sqrt{2}\frac{\delta}{1-\delta}c_3\right)^n, \tag{2.79}$$

where $g(x)$ is defined in (2.65). Consequently, by following passages similar to those in the proof of Lemma 2.1 in Appendix 2.6.1 and of Theorem 2.1 in Section 2.3.2, we conclude that

$$\mathbb{P}\left(Z_{n_0} \leq Z(W)\cdot 2^{-2n_0}\right) \geq C(W) - c_8\, 2^{-n_0/\mu}, \tag{2.80}$$

where c_8 is a constant. Note that in formula (2.42) $Z_{n_0+n_1}$ is upper bounded by a quantity that does not depend on x. In order to make this dependency appear, we use a procedure similar to that of the proof of Lemma 22 in [128]. As a result, we obtain that

$$\mathbb{P}\left(Z_{n_0+n_1} \leq x^{\frac{1}{2}\cdot 2^{\sum_{i=1}^{n_1} B_i}} \mid Z_{n_0} = x\right) \geq 1 - c_9\sqrt{x}(1-\log_2 x), \tag{2.81}$$

where c_9 is a constant. By combining (2.80) and (2.81), the result follows by using arguments similar to those of the proof of Theorem 2.3 in Section 2.4.2.

3 Scaling Exponent of List Decoding

Non usare mai il plurale majestatis.
Siamo convinti che faccia una pessima impressione.

Never use plurale majestatis. We believe it too pompous.

After developing, in the previous chapter, tight bounds on the scaling exponent of polar codes under SC decoding, we now investigate whether it is possible to improve such a scaling exponent by using a better decoding algorithm. Despite the excellent performance reported in [55], in this chapter[1], we provide some negative results for list decoding.

After reviewing some existing work in Section 3.1, we summarize our main contributions in Section 3.2: the scaling exponent does not change under MAP decoding with any finite list and, for the transmission over the BEC, also under genie-aided SC decoding for any finite number of helps from the genie. In Section 3.3, we discuss the result for MAP decoding with a list and, in Sections 3.4 and 3.5, we prove it for the special case of the BEC and for general BMS channels, respectively. In Section 3.6, we discuss the analysis for genie-aided SC decoding when the transmission takes place over the BEC. We defer some of the proofs of the intermediate lemmas to the appendix in Section 3.7.

3.1 Related Work

List decoding was introduced independently by Elias and Wozencraft [132,133] and it enables the receiver to collect L possible transmitted messages. An error is declared only if the correct message does not appear in the list.

The *error exponent* of list decoding schemes has been widely studied in the literature [134,135], and, for random coding, it has been proved that the introduction

[1] The material of this chapter is based on joint work with S. H. Hassani and R. Urbanke [130,131].

of a list with finite size L does not yield any change in this asymptotic regime, provided that the rate is close enough to capacity [120]. Improved bounds suitable for both random and structured linear block codes have been recently investigated [136].

As for the *scaling exponent*, for a random ensemble transmitted over a BEC with erasure probability ε, namely a BEC(ε), it can be shown that the error probability $P_{\mathrm{B}}(N, R, \varepsilon, L)$ scales as

$$P_{\mathrm{B}}(N, R, \varepsilon, L) \approx Q\bigg(\frac{\log_2 L}{\sqrt{N\varepsilon(1-\varepsilon)}} + \frac{\sqrt{N}(1-\varepsilon-R)}{\sqrt{\varepsilon(1-\varepsilon)}} \bigg), \tag{3.1}$$

where $Q(\cdot)$ is defined in (1.21).

To prove (3.1), consider a random matrix with NR rows and $N-E$ columns whose elements are i.i.d. random variables taking the values 0 and 1 with equal probability and where E is a binomial random variable with mean $N\varepsilon$ and variance $N\varepsilon(1-\varepsilon)$. Then, $P_{\mathrm{B}}(N, R, \varepsilon, L)$ is the probability that this matrix has rank $< NR - \log_2 L$. After some calculations and the application of Theorem 3.2.1 of [137], we obtain that the dominant term in $P_{\mathrm{B}}(N, R, \varepsilon, L)$ is given by $\mathbb{P}(E > N(1-R)+\log_2 L)$, which is expanded in (3.1).

As a result, the scaling exponent of random codes remains equal to 2 and even the mother curve[2] stays unchanged, namely $f(z) = Q(z/\sqrt{\varepsilon(1-\varepsilon)})$, for any $L \in \mathbb{N}$.

3.2 Main Results

The contributions of this chapter can be summarized as follows.

MAP decoding with a list. We show that the *scaling exponent does not improve* for any *finite list size*, for any BMS channel W, and for any family of linear codes whose minimum distance grows arbitrarily large when the block length N tends to infinity. By proving that the minimum distance of polar codes is unbounded in the limit $N \to +\infty$, we deduce that these conclusions also hold for polar codes. In particular, by means of a *Divide and Intersect* (DI) procedure, we show that the error probability of the MAP decoder with list size L, namely $P_{\mathrm{B}}^{\mathrm{MAP}}(N, R, W, L)$, is lower bounded by $P_{\mathrm{B}}^{\mathrm{MAP}}(N, R, W, L = 1)$ raised to an appropriate power times a suitable constant, both of which depend only on L. As a result, we see that list decoding has the potential of significantly improving the involved constants, but it does not change the scaling exponent.

Genie-aided SC decoding. Consider genie-aided SC decoding of polar codes for the transmission over the BEC. This decoder runs the SC algorithm and it can ask the value of a certain bit to the genie for a maximum of k times. The k-genie-aided SC decoder performs slightly worse than the SCL decoder with list size 2^k, but it is easier to analyze. We show that the *scaling exponent does not improve* for any *finite number of helps* from the genie. The proof technique is similar to that developed for MAP decoding and it is based on a DI bound.

[2] For a definition of the mother curve, see (2.2).

3.3 Analysis for MAP Decoding with a List

Let $\mathcal{C}_{\text{lin}}$ be a set of linear codes parameterized by their block length N and rate R. For each N and R, let $d_{\min}(N,R)$ denote the minimum distance. Consider the transmission over a BMS channel W with capacity $C(W) \in (0,1)$ and Bhattacharyya parameter $Z \in (0,1)$, defined in (1.3). Let $P_{\text{B}}^{\text{MAP}}(N,R,W,L)$ be the block error probability for the transmission over W under MAP decoding with list size L. In addition, denote by $\mathcal{C}_{\text{pol}}$ the set of polar codes when the transmission takes place over the BMS channel W.

The case of the BEC is handled separately. Indeed, for the transmission over an erasure channel, MAP decoding is reduced to solving a linear system over the finite field $\mathbb{F}_2$. Therefore, the number of codewords compatible with the received message is a power of 2. Let us assume that the MAP decoder with list size L declares an error if and only if the number of compatible codewords is strictly bigger than L. Consider a first MAP decoder with list size L_1, and a second MAP decoder whose list size L_2 is the biggest power of 2 smaller than L_1, i.e., $L_2 = 2^{\lfloor \log_2 L_1 \rfloor}$. Then, the performance of these two decoders are identical (the first one declares error if and only if the second one does). As a result, we can restrict our analysis to list sizes that are powers of 2, hence the bounds can be tightened. In addition, when dealing with a BEC, the proof is considerably simpler and it keeps the same flavor as the one valid for general channels.

3.3.1 Divide and Intersect (DI) Bounds

Theorem 3.1 (DI Bound - $\mathcal{C}_{\text{lin}}$)**.** *Consider the transmission using elements in $\mathcal{C}_{\text{lin}}$ over a BMS channel W with Bhattacharyya parameter Z and set $P_{\text{B}} \in (0,1)$. For any N and R so that*

$$P_{\text{B}}^{\text{MAP}}(N,R,W,L) > P_{\text{B}}, \tag{3.2}$$

$$d_{\min}(N,R) > \frac{\ln(P_{\text{B}}/8)}{\ln Z}, \tag{3.3}$$

the performance of the MAP decoder with list size $L+1$ ($2L$, if $W = BEC(\varepsilon)$) is lower bounded by

$$\begin{aligned} P_{\text{B}}^{\text{MAP}}(N,R,W,L+1) &\geq \frac{3}{16} \cdot \left(P_{\text{B}}^{\text{MAP}}(N,R,W,L)\right)^2, \\ P_{\text{B}}^{\text{MAP}}(N,R,\varepsilon,2L) &\geq \frac{3}{16} \cdot \left(P_{\text{B}}^{\text{MAP}}(N,R,\varepsilon,L)\right)^2. \end{aligned} \tag{3.4}$$

Theorem 3.2 (DI Bound - $\mathcal{C}_{\text{pol}}$)**.** *Consider the transmission using elements in $\mathcal{C}_{\text{pol}}$ over a BMS channel W. Fix $P_{\text{B}} \in (0,1)$ and pick any N such that*

$$N > 2^{\bar{n}(Z,C(W),P_{\text{B}})}, \tag{3.5}$$

where

$$\begin{aligned} \bar{n}(Z,C(W),P_{\text{B}}) &= 2\bar{m}(Z,P_{\text{B}}) - \ln(1-C(W)) \\ &\quad + \sqrt{-4\bar{m}(Z,P_{\text{B}}) \cdot \ln(1-C(W)) + (\ln(1-C(W)))^2}, \end{aligned} \tag{3.6}$$

with

$$\bar{m}(Z, P_{\mathrm{B}}) = \log_2\left(\frac{2\ln(P_{\mathrm{B}}/8)\cdot\ln(1-Z)}{\ln Z\cdot\ln\left(1-Z^{\frac{4\ln(P_{\mathrm{B}}/8)}{\ln Z}}\right)}\right), \tag{3.7}$$

and any sufficiently large R so that

$$P_{\mathrm{B}}^{\mathrm{MAP}}(N,R,W,L) > P_{\mathrm{B}}. \tag{3.8}$$

Then, the bounds (3.4) *hold.*

The corollary below follows by induction.

Corollary 3.1 (DI Bound - Any L)**.** *Consider the transmission using elements in $\mathcal{C}_{\mathrm{lin}}$ over a BMS channel W. Fix $P_{\mathrm{B}} \in (0,1)$ and define the following recursion,*

$$P_{\mathrm{B}}(m+1) = \frac{3}{16}\big(P_{\mathrm{B}}(m)\big)^2, \qquad m \in \mathbb{N}\,, \tag{3.9}$$

with the initial condition $P_{\mathrm{B}}(1) = P_{\mathrm{B}}$. Pick any N and R such that (3.2) *and* (3.3) *hold with $P_{\mathrm{B}}(L)$ instead of P_{B}, or, if the code is in $\mathcal{C}_{\mathrm{pol}}$, any N satisfying* (3.5) *and any sufficiently large R satisfying* (3.8) *with $P_{\mathrm{B}}(L)$ instead of P_{B}. Then, the performance of the MAP decoder with list size $L+1$ is lower bounded by*

$$P_{\mathrm{B}}^{\mathrm{MAP}}(N,R,W,L+1) \geq \left(\frac{3}{16}\right)^{2^L-1}\cdot\big(P_{\mathrm{B}}^{\mathrm{MAP}}(N,R,W,L=1)\big)^{2^L}. \tag{3.10}$$

If $W = BEC(\varepsilon)$, consider the recursion (3.9) *with the initial condition $P_{\mathrm{B}}(0) = P_{\mathrm{B}}$. If* (3.2)-(3.3) *and* (3.5)-(3.8) *are satisfied with $P_{\mathrm{B}}(\log_2 L)$ instead of P_{B} for codes in $\mathcal{C}_{\mathrm{lin}}$ and $\mathcal{C}_{\mathrm{pol}}$, respectively, then the performance of the MAP decoder with list size $2L$ is lower bounded by*

$$P_{\mathrm{B}}^{\mathrm{MAP}}(N,R,\varepsilon,2L) \geq \left(\frac{3}{16}\right)^{2L-1}\cdot\big(P_{\mathrm{B}}^{\mathrm{MAP}}(N,R,\varepsilon,L=1)\big)^{2L}. \tag{3.11}$$

3.3.2 Scaling Exponent

An immediate consequence of the DI bounds is that the scaling exponent does not change as long as L is fixed and finite. The theorem below bounds the scaling behavior of the MAP decoder with any finite list size L, and its proof is easily deduced from Corollary 3.1.

Theorem 3.3 (Scaling Exponent - MAP Decoding with a List)**.** *Consider the set of polar codes $\mathcal{C}_{\mathrm{pol}}$ transmitted over a BMS channel W. Assume that the limit* (2.2) *exists under MAP decoding, i.e.,*

$$\lim_{N\to\infty:\ N^{1/\mu}(C(W)-R)=z} P_{\mathrm{B}}^{\mathrm{MAP}}(N,R,W) = f(z),$$

where μ is the scaling exponent and f the mother curve. Then, for any $L \in \mathbb{N}$,

$$\limsup_{N\to\infty:\ N^{1/\mu}(C(W)-R)=z} P_{\mathrm{B}}^{\mathrm{MAP}}(N,R,W,L) \leq f(z), \tag{3.12}$$

$$\liminf_{N\to\infty:\ N^{1/\mu}(C(W)-R)=z} P_{\mathrm{B}}^{\mathrm{MAP}}(N,R,W,L) \geq \left(\frac{3}{16}\right)^{2^{L-1}-1}\cdot\big(f(z)\big)^{2^{L-1}}. \tag{3.13}$$

In words, if a scaling law holds for the MAP decoder with list size L, the scaling exponent μ is the same as that for the original MAP decoder without list. Therefore, the speed at which capacity is approached as the block length grows large does not depend on the list size, provided that L remains fixed. Notice that, in general, Theorem 3.3 holds for any set of linear codes whose minimum distance is unbounded as the block length grows large.

3.4 Proof of DI Bounds for BEC

As the name suggests, the DI procedure has two main ingredients: the *Intersect* step is based on the correlation inequality stated in Section 3.4.1; the *Divide* step is based on the existence of a suitable subset of codewords, which is discussed in Section 3.4.2. We prove the bound for the simple case $L = 1$ for linear and polar codes in Sections 3.4.3 and 3.4.4, respectively. We present the generalization to any list size in Section 3.4.5.

3.4.1 Intersect Step: Correlation Inequality

As the BEC is a symmetric channel, we can assume that the all-zero codeword has been transmitted. As the BEC does not introduce errors, the MAP decoder outputs all the codewords compatible with the received message. An error is declared if and only if the all-zero codeword is not the unique candidate, i.e., there is more than one candidate codeword.

Let us map the channel output into the erasure pattern $y = (y_1, \cdots, y_N) \in \{0,1\}^N$, with $y_i = 1$ meaning that the i-th BEC has yielded an erasure symbol and $y_i = 0$, otherwise. Let G_y be the part of the generator matrix G obtained by eliminating the columns corresponding to the erased symbols, i.e., all the columns of index i such that $y_i = 1$. It is easy to check that the MAP decoder outputs the information vector $u = (u_1, \cdots, u_{NR})$ if and only if $uG_y = 0$. Define E_u to be the set of all the erasure patterns such that u solves $uG_y = 0$, i.e.,

$$E_u = \{y \in \{0,1\}^N \mid uG_y = 0\}. \tag{3.14}$$

Let I_u be the set of positions i in which $(uG)_i$ equals 1, namely,

$$I_u = \{i \in [N] \mid (uG)_i = 1\}. \tag{3.15}$$

Since $P_{\mathrm{B}}^{\mathrm{MAP}}(N, R, \varepsilon, L = 1)$ is the probability that there exists a non-zero informative vector u that satisfies $uG_y = 0$, we have

$$\mathbb{P}(\bigcup_{u \in U} E_u) = P_{\mathrm{B}}^{\mathrm{MAP}}(N, R, \varepsilon, L = 1), \tag{3.16}$$

with

$$U = \mathbb{F}_2^{NR} \setminus 0_{1:NR}, \tag{3.17}$$

where $0_{1:NR}$ denotes a sequence of NR 0s.

We start with two simple lemmas that compute $\mathbb{P}(E_u)$ and that show the positive correlation between the events (3.14).

Lemma 3.1. *Let $u \in \mathbb{F}_2^{NR}$ and let E_u be defined in* (3.14)*. Then,*

$$\mathbb{P}(E_u) = \varepsilon^{|I_u|}, \tag{3.18}$$

where I_u is given by (3.15)*.*

Proof. Observe that u solves $uG_y = 0$ if and only if all the positions i such that $(uG)_i = 1$ are erased by the BEC(ε). Therefore, $\mathbb{P}(E_u)$ equals the probability that $|I_u|$ independent erasures at those positions occur, which implies (3.18). □

Lemma 3.2. *Let $u, \tilde{u} \in \mathbb{F}_2^{NR}$. Then,*

$$\mathbb{P}(E_u \cap E_{\tilde{u}}) \geq \mathbb{P}(E_u) \cdot \mathbb{P}(E_{\tilde{u}}). \tag{3.19}$$

Proof. By definition (3.15), we obtain

$$\mathbb{P}(E_u \cap E_{\tilde{u}}) = \varepsilon^{|I_u \cup I_{\tilde{u}}|} = \varepsilon^{|I_u|+|I_{\tilde{u}}|-|I_u \cap I_{\tilde{u}}|} \geq \varepsilon^{|I_u|+|I_{\tilde{u}}|} = \mathbb{P}(E_u) \cdot \mathbb{P}(E_{\tilde{u}}), \tag{3.20}$$

which gives (3.19). □

Let us now generalize Lemma 3.2 to unions of sets.

Lemma 3.3 (Positive Correlation - BEC(ε), $L = 1$)**.** *Let $U_1, U_2 \subset \mathbb{F}_2^{NR}$. Then,*

$$\mathbb{P}\Big(\bigcup_{u \in U_1} E_u \cap \bigcup_{\tilde{u} \in U_2} E_{\tilde{u}}\Big) \geq \mathbb{P}\Big(\bigcup_{u \in U_1} E_u\Big) \cdot \mathbb{P}\Big(\bigcup_{\tilde{u} \in U_2} E_{\tilde{u}}\Big). \tag{3.21}$$

The proof of this result can be found in Appendix 3.7.1 and comes from an application of the *FKG inequality*, originally proposed in [138].

3.4.2 Divide Step: Existence of a Suitable Subset of Codewords

The purpose of this part is to show that there exists $U_1 \subset U$ such that $\mathbb{P}(\bigcup_{u \in U_1} E_u)$ is slightly smaller than $\frac{1}{2} P_{\mathrm{B}}^{\mathrm{MAP}}(N, R, \varepsilon, L = 1)$. To do so, we first upper bound $\mathbb{P}(E_u)$ for all $u \in U$.

Lemma 3.4 (No Big Jumps - BEC(ε))**.** *Let $P_{\mathrm{B}} \in (0, 1)$ and $\varepsilon \in (0, 1)$. Then, for any N and R so that*

$$d_{\min}(N, R) > \frac{\ln(P_{\mathrm{B}}/8)}{\ln \varepsilon}, \tag{3.22}$$

the probability of E_u is bounded by

$$\mathbb{P}(E_u) < \frac{P_{\mathrm{B}}}{8}, \qquad \forall\ u \in U. \tag{3.23}$$

Proof. From Lemma 3.1 and the definition of minimum distance, we obtain that

$$\mathbb{P}(E_u) = \varepsilon^{|I_u|} \leq \varepsilon^{d_{\min}}. \tag{3.24}$$

Using (3.22), the thesis follows. □

The existence of a subset of codewords with the desired property is an immediate consequence of the previous lemma.

Corollary 3.2 (Existence of U_1)**.** *Let $P_{\mathrm{B}} \in (0,1)$ and $\varepsilon \in (0,1)$. Then, for any N and R so that (3.22) and $P_{\mathrm{B}}^{\mathrm{MAP}}(N,R,\varepsilon,L=1) > P_{\mathrm{B}}$ hold, there exists $U_1 \subset U$ that satisfies*

$$\begin{aligned}\mathbb{P}(\bigcup_{u\in U_1} E_u) &\geq \frac{3}{8} P_{\mathrm{B}}^{\mathrm{MAP}}(N,R,\varepsilon,L=1),\\ \mathbb{P}(\bigcup_{u\in U_1} E_u) &\leq \frac{1}{2} P_{\mathrm{B}}^{\mathrm{MAP}}(N,R,\varepsilon,L=1).\end{aligned} \tag{3.25}$$

3.4.3 DI Bound for Linear Codes

At this point, we are ready to present the proof of Theorem 3.1 for the BEC and for a list size $L=1$. Recall that the Bhattacharyya parameter of a BEC(ε) is $Z=\varepsilon$.

Proof of Theorem 3.1 for BEC(ε), $L=1$. Pick U_1 that satisfies (3.25) and let $U_2 = U \setminus U_1$. Consequently,

$$\begin{aligned}\frac{3}{8}\cdot P_{\mathrm{B}}^{\mathrm{MAP}}(N,R,\varepsilon,L=1) &\leq \mathbb{P}(\bigcup_{u\in U_1} E_u),\\ \frac{1}{2}\cdot P_{\mathrm{B}}^{\mathrm{MAP}}(N,R,\varepsilon,L=1) &\leq \mathbb{P}(\bigcup_{\tilde{u}\in U_2} E_{\tilde{u}}).\end{aligned}$$

Hence,

$$\frac{3}{16}\cdot \left(P_{\mathrm{B}}^{\mathrm{MAP}}(N,R,\varepsilon,L=1)\right)^2 \leq \mathbb{P}(\bigcup_{u\in U_1} E_u)\cdot \mathbb{P}(\bigcup_{\tilde{u}\in U_2} E_{\tilde{u}}).$$

In addition, the following chain of inequalities holds,

$$\begin{aligned}&\mathbb{P}(\bigcup_{u\in U_1} E_u)\cdot \mathbb{P}(\bigcup_{\tilde{u}\in U_2} E_{\tilde{u}}) \leq \mathbb{P}(\bigcup_{u\in U_1} E_u \cap \bigcup_{\tilde{u}\in U_2} E_{\tilde{u}})\\ &= \mathbb{P}(\bigcup_{u\in U_1,\tilde{u}\in U_2} E_u \cap E_{\tilde{u}}) \leq \mathbb{P}(\bigcup_{u,\tilde{u}\in U, u\neq\tilde{u}} E_u \cap E_{\tilde{u}}),\end{aligned}$$

where the first inequality comes from the application of Lemma 3.3 and the last passage is a direct consequence of $U_1 \cap U_2 = \emptyset$. Noticing that

$$\mathbb{P}(\bigcup_{u,\tilde{u}\in U, u\neq\tilde{u}} E_u \cap E_{\tilde{u}}) = P_{\mathrm{B}}^{\mathrm{MAP}}(N,R,\varepsilon,L=2),$$

we obtain the desired result. □

3.4.4 DI Bound for Polar Codes

In order to apply the bound to polar codes, it suffices to prove the lower bound on the minimum distance, as required in (3.22).

Lemma 3.5 ($d_{\min}$ of Polar Codes - BEC(ε))**.** *Consider a polar code in $\mathcal{C}_{\mathrm{pol}}$ for a BEC(ε). Let $P_{\mathrm{B}} \in (0,1)$, $\varepsilon \in (0,1)$, and $N > 2^{\bar{n}(\varepsilon,P_{\mathrm{B}})}$, where*

$$\bar{n}(\varepsilon,P_{\mathrm{B}}) = 2\bar{m}(\varepsilon,P_{\mathrm{B}}) - \ln\varepsilon + \sqrt{-4\bar{m}(\varepsilon,P_{\mathrm{B}})\cdot\ln\varepsilon + (\ln\varepsilon)^2}, \tag{3.26}$$

with

$$\bar{m}(\varepsilon, P_{\mathrm{B}}) = \log_2\left(\frac{2\ln(P_{\mathrm{B}}/8)\cdot\ln(1-\varepsilon)}{\ln\varepsilon\cdot\ln\left(1-\varepsilon^{\frac{2\ln(P_{\mathrm{B}}/8)}{\ln\varepsilon}}\right)}\right). \tag{3.27}$$

Then, the lower bound on $d_{\min}$ (3.22) *holds.*

The proof of Lemma 3.5 is in Appendix 3.7.2. Due to this result, Theorem 3.2 follows from Theorem 3.1. Comparing (3.7) with (3.27), we notice that, for the BEC, the constraint on N is less tight than the one required for any BMS channel.

3.4.5 Generalization to Any List Size

Set $l = \log_2 L$ and define $E_{\mathrm{sp}(u^{(1)},\cdots,u^{(l)})}$ to be the set of all erasure patterns y such that the set of solutions of the linear system $uG_y = 0$ contains the linear span generated by $\{u^{(1)},\cdots,u^{(l)}\}$, i.e.,

$$\begin{aligned} E_{\mathrm{sp}(u^{(1)},\cdots,u^{(l)})} &= \bigcap_{u\in\mathrm{span}(u^{(1)},\cdots,u^{(l)})} E_u \\ &= \{y\in\{0,1\}^N \mid uG_y = 0 \quad \forall\ u\in\mathrm{span}(u^{(1)},\cdots,u^{(l)})\}. \end{aligned} \tag{3.28}$$

Consider the set LS_l containing all the linear spans of $\mathbb{F}_2^{NR}$ with 2^l elements. In formulae,

$$\mathrm{LS}_l = \{\mathrm{span}(u^{(1)},\cdots,u^{(l)}) \mid u^{(i)}\in\mathbb{F}_2^{NR}\ \forall i\in[l], |\mathrm{span}(u^{(1)},\cdots,u^{(l)})| = 2^l\}. \tag{3.29}$$

Since $P_{\mathrm{B}}^{\mathrm{MAP}}(N,R,\varepsilon,L)$ is the probability that the solutions to the linear system $uG_y = 0$ form a linear span of cardinality strictly greater than L, we have

$$\mathbb{P}\Big(\bigcup_{\mathrm{span}(u^{(1)},\cdots,u^{(l+1)})\in\mathrm{LS}_{l+1}} E_{\mathrm{sp}(u^{(1)},\cdots,u^{(l+1)})}\Big) = P_{\mathrm{B}}^{\mathrm{MAP}}(N,R,\varepsilon,L). \tag{3.30}$$

For the *Intersect* step, we need now the generalization of Lemma 3.3, which is contained in Lemma 3.6.

Lemma 3.6 (Positive Correlation - BEC(ε), Any L)**.** *Let* $P_1, P_2 \subset \mathrm{LS}_l$*. Then,*

$$\begin{aligned} &\mathbb{P}\Big(\bigcup_{\mathrm{span}(u^{(1)},\cdots,u^{(l)})\in P_1} E_{\mathrm{sp}(u^{(1)},\cdots,u^{(l)})} \cap \bigcup_{\mathrm{span}(\tilde{u}^{(1)},\cdots,\tilde{u}^{(l)})\in P_2} E_{\mathrm{sp}(\tilde{u}^{(1)},\cdots,\tilde{u}^{(l)})}\Big) \\ &\geq \mathbb{P}\Big(\bigcup_{\mathrm{span}(u^{(1)},\cdots,u^{(l)})\in P_1} E_{\mathrm{sp}(u^{(1)},\cdots,u^{(l)})}\Big)\cdot\mathbb{P}\Big(\bigcup_{\mathrm{span}(\tilde{u}^{(1)},\cdots,\tilde{u}^{(l)})\in P_2} E_{\mathrm{sp}(\tilde{u}^{(1)},\cdots,\tilde{u}^{(l)})}\Big). \end{aligned} \tag{3.31}$$

The proof of Lemma 3.6 is given in Appendix 3.7.3. We are going to need also the subsequent simple result concerning the intersection of events (3.28).

Lemma 3.7. *For any* $\mathrm{span}(u^{(1)},\cdots,u^{(l)})$ *and* $\mathrm{span}(\tilde{u}^{(1)},\cdots,\tilde{u}^{(l)})$*,*

$$E_{\mathrm{sp}(u^{(1)},\cdots,u^{(l)})} \cap E_{\mathrm{sp}(\tilde{u}^{(1)},\cdots,\tilde{u}^{(l)})} = E_{\mathrm{sp}(u^{(1)},\cdots,u^{(l)},\tilde{u}^{(1)},\cdots,\tilde{u}^{(l)})}. \tag{3.32}$$

Proof. Note that

$$\mathrm{span}(u^{(1)},\cdots,u^{(l)},\tilde{u}^{(1)},\cdots,\tilde{u}^{(l)}) \supset \mathrm{span}(u^{(1)},\cdots,u^{(l)}) \cup \mathrm{span}(\tilde{u}^{(1)},\cdots,\tilde{u}^{(l)}).$$

Then,

$$E_{\mathrm{sp}(u^{(1)},\cdots,u^{(l)})} \cap E_{\mathrm{sp}(\tilde{u}^{(1)},\cdots,\tilde{u}^{(l)})} \supset E_{\mathrm{sp}(u^{(1)},\cdots,u^{(l)},\tilde{u}^{(1)},\cdots,\tilde{u}^{(l)})}.$$

On the contrary, by linearity of the code, for any $u \in \mathrm{span}(u^{(1)},\cdots,u^{(l)})$ and any $v \in \mathrm{span}(\tilde{u}^{(1)},\cdots,\tilde{u}^{(l)})$, if $uG_y = 0$ and $vG_y = 0$, then $wG_y = 0$ for all $w \in \{u+v : u \in \mathrm{span}(u^{(1)},\cdots,u^{(l)}), v \in \mathrm{span}(\tilde{u}^{(1)},\cdots,\tilde{u}^{(l)})\}$. As a result,

$$E_{\mathrm{sp}(u^{(1)},\cdots,u^{(l)})} \cap E_{\mathrm{sp}(\tilde{u}^{(1)},\cdots,\tilde{u}^{(l)})} \subset E_{\mathrm{sp}(u^{(1)},\cdots,u^{(l)},\tilde{u}^{(1)},\cdots,\tilde{u}^{(l)})},$$

and the thesis follows. □

For the *Divide* step, Corollary 3.3 generalizes the result of Corollary 3.2 to any list size L.

Corollary 3.3 (Existence of P_1)**.** *Let $P_{\mathrm{B}} \in (0,1)$ and $\varepsilon \in (0,1)$. Then, for any R and N satisfying*

$$P_{\mathrm{B}}^{\mathrm{MAP}}(N,R,\varepsilon,L) > P_{\mathrm{B}}, \tag{3.33}$$

$$d_{\min} > \frac{\ln(P_{\mathrm{B}}/8)}{\ln\varepsilon}, \tag{3.34}$$

there exists $P_1 \subset \mathrm{LS}_{l+1}$ such that

$$\begin{aligned} \mathbb{P}\Big(\bigcup_{\mathrm{span}(u^{(1)},\cdots,u^{(l+1)})\in P_1} E_{\mathrm{sp}(u^{(1)},\cdots,u^{(l+1)})}\Big) &\geq \frac{3}{8}P_{\mathrm{B}}^{\mathrm{MAP}}(N,R,\varepsilon,L), \\ \mathbb{P}\Big(\bigcup_{\mathrm{span}(u^{(1)},\cdots,u^{(l+1)})\in P_1} E_{\mathrm{sp}(u^{(1)},\cdots,u^{(l+1)})}\Big) &\leq \frac{1}{2}P_{\mathrm{B}}^{\mathrm{MAP}}(N,R,\varepsilon,L). \end{aligned} \tag{3.35}$$

At this point, we can prove Theorem 3.1 for the BEC and for any list size L.

Proof of Theorem 3.1 for BEC(ε), any L. Pick P_1 that satisfies (3.35) and let $P_2 = \mathrm{LS}_{l+1} \setminus P_1$. Consequently, applying Lemma 3.6 and 3.7, we have

$$\begin{aligned} &\frac{3}{16}\cdot\big(P_{\mathrm{B}}^{\mathrm{MAP}}(N,R,\varepsilon,L)\big)^2 \\ &\quad\leq \mathbb{P}\Big(\bigcup_{\mathrm{span}(u^{(1)},\cdots,u^{(l+1)})\in P_1} E_{\mathrm{sp}(u^{(1)},\cdots,u^{(l+1)})}\Big)\cdot\mathbb{P}\Big(\bigcup_{\mathrm{span}(\tilde{u}^{(1)},\cdots,\tilde{u}^{(l+1)})\in P_2} E_{\mathrm{sp}(\tilde{u}^{(1)},\cdots,\tilde{u}^{(l+1)})}\Big) \\ &\quad\leq \mathbb{P}\Big(\bigcup_{\mathrm{span}(u^{(1)},\cdots,u^{(l+1)})\in P_1} E_{\mathrm{sp}(u^{(1)},\cdots,u^{(l+1)})} \cap \bigcup_{\mathrm{span}(\tilde{u}^{(1)},\cdots,\tilde{u}^{(l+1)})\in P_2} E_{\mathrm{sp}(\tilde{u}^{(1)},\cdots,\tilde{u}^{(l+1)})}\Big) \\ &\quad= \mathbb{P}\Big(\bigcup_{\substack{\mathrm{span}(u^{(1)},\cdots,u^{(l+1)})\in P_1 \\ \mathrm{span}(\tilde{u}^{(1)},\cdots,\tilde{u}^{(l+1)})\in P_2}} E_{\mathrm{sp}(u^{(1)},\cdots,u^{(l+1)},\tilde{u}^{(1)},\cdots,\tilde{u}^{(l+1)})}\Big) \\ &\quad\leq P_{\mathrm{B}}^{\mathrm{MAP}}(N,R,\varepsilon,2L), \end{aligned}$$

where the last inequality is due to the fact that

$$|\mathrm{span}(u^{(1)},\cdots,u^{(l+1)},\tilde{u}^{(1)},\cdots,\tilde{u}^{(l+1)})| \geq 2^{l+2} = 4L > 2L,$$

as $P_1 \cap P_2 = \emptyset$. □

3.5 Proof of DI Bounds for BMS Channels

3.5.1 Case $L = 1$

Since the information vectors are equiprobable, the MAP decision rule is given by

$$\hat{u} = \arg\max_{\tilde{u}} p(y|\tilde{u}).$$

Define E'_u as the set of all y such that $p(y|u) \geq p(y|0_{1:NR})$. Simple algebraic manipulations show that

$$E'_u = \{y \in \mathcal{Y}^N \mid \sum_{i=1}^{N} \ln \frac{p(y_i|(uG)_i)}{p(y_i|0)} \geq 0\} = \{y \in \mathcal{Y}^N \mid \sum_{i \in I_u} \ln \frac{p(y_i|1)}{p(y_i|0)} \geq 0\}, \quad (3.36)$$

where $\mathcal{Y}$ is the output alphabet of the channel and I_u is defined in (3.15).

Note that $P_{\mathrm{B}}^{\mathrm{MAP}}(N, R, \varepsilon, L = 1)$ is the probability that there exists a non-zero informative vector u such that $p(y|u) \geq p(y|0_{1:NR})$. Then, we have

$$\mathbb{P}(\bigcup_{u \in U} E'_u) = P_{\mathrm{B}}^{\mathrm{MAP}}(N, R, W, L = 1). \quad (3.37)$$

For the *Intersect* step, we generalize the inequality of Lemma 3.3 with the correlation result of Lemma 3.8.

Lemma 3.8 (Positive Correlation - BMS Channels, $L = 1$)**.** *Let $U'_1, U'_2 \subset \mathbb{F}_2^{NR}$. Then,*

$$\mathbb{P}(\bigcup_{u \in U'_1} E'_u \cap \bigcup_{\tilde{u} \in U'_2} E'_{\tilde{u}}) \geq \mathbb{P}(\bigcup_{u \in U'_1} E'_u) \cdot \mathbb{P}(\bigcup_{\tilde{u} \in U'_2} E'_{\tilde{u}}). \quad (3.38)$$

The proof of Lemma 3.8 can be found in Appendix 3.7.4.

For the *Divide* step, we need to show that $\mathbb{P}(E'_u)$ can be made as small as we want, as done in Lemma 3.4 for the events (3.14). This result is provided by Lemma 3.9, stated and proven below.

Lemma 3.9 (No Big Jumps - BMS Channels)**.** *Let $P_{\mathrm{B}} \in (0, 1)$ and $Z \in (0, 1)$. Then, for any N and R so that*

$$d_{\min}(N, R) > \frac{\ln(P_{\mathrm{B}}/8)}{\ln Z}, \quad (3.39)$$

the probability of E'_u can be bounded as

$$\mathbb{P}(E'_u) < \frac{P_{\mathrm{B}}}{8}, \qquad \forall\ u \in U. \quad (3.40)$$

Proof. By applying Lemma 4.66 of [44], it is possible to relate the probability of E'_u and the Bhattacharyya parameter Z of the BMS channel W as

$$\mathbb{P}(E'_u) \leq Z^{|I_u|}. \quad (3.41)$$

Since $|I_u| \geq d_{\min} > \ln(P_{\mathrm{B}}/8)/\ln Z$, the thesis easily follows. □

From Lemma 3.9 we deduce a result similar to that of Corollary 3.2. Then, by using also Lemma 3.8 and by following the same procedure seen at the end of Section 3.4.3, the proof of Theorem 3.1 with $L = 1$ for any BMS channel is readily obtained.

Lemma 3.10 generalizes the result of Lemma 3.5, showing that for N big enough the required lower bound on the minimum distance holds. Hence, the DI bound and the subsequent scaling result are true for sequences of polar codes.

Lemma 3.10 ($d_{\min}$ of Polar Codes - BMS Channels)**.** *Let $P_{\mathrm{B}} \in (0,1)$, $Z \in (0,1)$, and $N > 2^{\bar{n}(Z,C(W),P_{\mathrm{B}})}$, where $\bar{n}(Z, C(W), P_{\mathrm{B}})$ is given by* (3.6)*. Then, the lower bound on $d_{\min}$* (3.39) *holds.*

The proof of Lemma 3.10 is in Appendix 3.7.5.

3.5.2 Generalization to Any List Size

Denote by $E'_{u^{(1)},\cdots,u^{(L)}}$ be the set of all y such that $p(y|u) \geq p(y|0_{1:NR})$ for all $u \in \{u^{(1)}, \cdots, u^{(L)}\}$, i.e.,

$$\begin{aligned} E'_{u^{(1)},\cdots,u^{(L)}} &= \bigcap_{u \in \{u^{(1)},\cdots,u^{(L)}\}} E'_u = \{y \in \mathcal{Y}^N \mid \\ &\sum_{i=1}^{N} \ln \frac{p(y_i|(uG)_i)}{p(y_i|0)} \geq 0 \quad \forall\ u \in \{u^{(1)}, \cdots, u^{(L)}\}\}. \end{aligned} \tag{3.42}$$

Consider the set SS_L containing all the subsets of L distinct elements of $\mathbb{F}_2^{NR}$. In formulae,

$$\mathrm{SS}_L = \{\{u^{(1)}, \cdots, u^{(L)}\} : u^{(i)} \in \mathbb{F}_2^{NR}\ \forall i \in [L], u^{(i)} \neq u^{(j)}\ \ \forall\ i \neq j\}. \tag{3.43}$$

Since $P_{\mathrm{B}}^{\mathrm{MAP}}(N, R, W, L)$ is the probability that there are at least L distinct non-zero information vectors $u^{(1)}, \cdots, u^{(L)}$ such that $p(y|u) \geq p(y|0_{1:NR})$ for all $u \in \{u^{(1)}, \cdots, u^{(L)}\}$, we have

$$\mathbb{P}\Big(\bigcup_{\{u^{(1)},\cdots,u^{(L)}\} \in \mathrm{SS}_L} E'_{u^{(1)},\cdots,u^{(L)}} \Big) = P_{\mathrm{B}}^{\mathrm{MAP}}(N, R, W, L). \tag{3.44}$$

In order to prove Theorem 3.1 for any fixed list size L and for any BMS channel, we can follow similar steps to those of Section 3.4.5 and the result is readily obtained. The only part that requires some further investigation consists in the generalization of Lemma 3.8 with the result below, which is proved in Appendix 3.7.6.

Lemma 3.11 (Positive Correlation - BMS Channel, any L)**.** *Let $P'_1, P'_2 \subset \mathrm{SS}_L$. Then,*

$$\begin{aligned} &\mathbb{P}\Big(\bigcup_{\{u^{(1)},\cdots,u^{(L)}\} \in P'_1} E'_{u^{(1)},\cdots,u^{(L)}} \cap \bigcup_{\{\tilde{u}^{(1)},\cdots,\tilde{u}^{(L)}\} \in P'_2} E'_{\tilde{u}^{(1)},\cdots,\tilde{u}^{(L)}} \Big) \\ &\geq \mathbb{P}\Big(\bigcup_{\{u^{(1)},\cdots,u^{(L)}\} \in P'_1} E'_{u^{(1)},\cdots,u^{(L)}} \Big) \cdot \mathbb{P}\Big(\bigcup_{\{\tilde{u}^{(1)},\cdots,\tilde{u}^{(L)}\} \in P'_2} E'_{\tilde{u}^{(1)},\cdots,\tilde{u}^{(L)}} \Big). \end{aligned} \tag{3.45}$$

3.6 Analysis for Genie-Aided SC Decoding

Consider a polar code in $\mathcal{C}_{\text{pol}}$ transmitted over a BEC(ε). In accordance with the notation defined in Section 1.3, let $n = \log_2 N$ and denote by $W_n^{(i)}$ the i-th synthetic channel, which is a BEC of Bhattacharyya parameter $Z_n^{(i)}$. Let $P_{\text{B}}^{\text{SC}}(N, R, \varepsilon, k)$ be the block error probability under SC decoding aided by a k-genie. More precisely, a k-genie-aided SC decoder runs the usual SC algorithm with the following difference: when we reach a synthetic channel associated with an information bit that is erased, i.e., we cannot decide on the value of a certain information bit, the genie tells the value of the erased bit, and it does so a maximum of k times. An error is declared if and only if the decoder requires more than k helps from the genie.

Consider now SCL decoding for the transmission over the BEC. The SCL decoder also runs the usual SC algorithm and, when we reach a synthetic channel associated with an information bit that is erased, say $W_n^{(i)}$, it takes into account both the possibilities, namely $u_i = 0$ and $u_i = 1$, and it lets the two decoding paths evolve independently. In addition, when we reach a synthetic channel associated with a frozen bit, the SCL decoder gains new information, i.e., it learns the value of the linear combination of some of the previous bits. Therefore, some of the existing decoding paths might not be compatible with this extra information and they are removed from the list. An error is declared if and only if at any point during the decoding process the decoder requires to store more than L decoding paths.

Note that a k-genie-aided SC decoder and an SCL decoder with list size 2^k behave similarly but not identically. When we reach a synthetic channel associated with an information bit that is erased, the former uses one of the helps from the genie, and the latter doubles the number of decoding paths. However, when we reach a synthetic channel associated with a frozen bit, the SCL decoder can reduce the number of decoding paths, whereas it is not possible to gain new helps from the genie. Therefore, the SCL decoder always succeeds when the genie-aided decoder succeeds, but it might also succeed in some cases where the genie-aided decoder fails.

3.6.1 DI Bound and Scaling Exponent

Theorem 3.4 (DI Bound - Genie-Aided Decoding)**.** *Consider the transmission of a polar code in $\mathcal{C}_{\text{pol}}$ over a BEC(ε) and fix $P_{\text{B}} \in (0, 1)$. Pick N big enough and any R that ensures*

$$\frac{1}{4} > P_{\text{B}}^{\text{SC}}(N, R, \varepsilon, k) > P_{\text{B}}. \tag{3.46}$$

Then, the performance of the $k+1$-genie-aided SC decoder is lower bounded by

$$P_{\text{B}}^{\text{SC}}(N, R, \varepsilon, k+1) \geq \frac{3}{16} \cdot \left(P_{\text{B}}^{\text{SC}}(N, R, \varepsilon, k)\right)^2. \tag{3.47}$$

By induction, the corollary below easily follows.

Corollary 3.4 (DI Bound - Genie-Aided Decoding, Any k)**.** *Consider the transmission of a polar code in $\mathcal{C}_{\text{pol}}$ over a BEC(ε). Fix $P_{\text{B}} \in (0, 1)$ and consider the recursion* (3.9) *with the initial condition $P_{\text{B}}(0) = P_{\text{B}}$. Pick N big enough and R*

such that (3.46) *holds with* $P_{\mathrm{B}}(k)$ *instead of* P_{B}. *Then, the performance of the* $k+1$*-genie-aided SC decoder is lower bounded by*

$$P_{\mathrm{B}}^{\mathrm{SC}}(N,R,\varepsilon,k+1) \geq \left(\frac{3}{16}\right)^{2^{k+1}-1} \cdot \left(P_{\mathrm{B}}^{\mathrm{SC}}(N,R,\varepsilon,k=0)\right)^{2^{k+1}}. \tag{3.48}$$

Roughly speaking, Theorem 3.4 implies that the scaling exponent cannot change under SC decoding for any fixed number of helps from the genie. This statement can be formalized as follows.

Theorem 3.5 (Scaling Exponent - Genie-Aided Decoding)**.** *Consider the set of polar codes* $\mathcal{C}_{\mathrm{pol}}$ *transmitted over a BEC*(ε). *Assume that the limit* (2.2) *exists under SC decoding, i.e.,*

$$\lim_{N\to\infty:\ N^{1/\mu}(C(W)-R)=z} P_{\mathrm{B}}^{\mathrm{SC}}(N,R,W) = f(z),$$

where μ *is the scaling exponent and* f *the mother curve. Then, for any* $k \in \mathbb{N}$,

$$\limsup_{N\to\infty:\ N^{1/\mu}(C(W)-R)=z} P_{\mathrm{B}}^{\mathrm{SC}}(N,R,\varepsilon,k) \leq f(z), \tag{3.49}$$

$$\liminf_{N\to\infty:\ N^{1/\mu}(C(W)-R)=z} P_{\mathrm{B}}^{\mathrm{MAP}}(N,R,\varepsilon,k) \geq \left(\frac{3}{16}\right)^{2^{k}-1} \cdot \left(f(z)\right)^{2^{k}}. \tag{3.50}$$

3.6.2 Proof of DI Bound

Let $y \in \{0,1\}^N$ denote the erasure pattern of the channel and, for $i \in \{1,\cdots,N\}$, let F_i be the set containing all y such that $W_n^{(i)}$ erases, i.e.,

$$F_i = \{y \in \{0,1\}^N \mid W_n^{(i)} \text{ erases}\}. \tag{3.51}$$

Denoting by $\mathcal{F}^c$ the set of unfrozen positions, it is clear that

$$\mathbb{P}\Big(\bigcup_{i\in\mathcal{F}^c} F_i\Big) = P_{\mathrm{B}}^{\mathrm{SC}}(N,R,\varepsilon,k=0). \tag{3.52}$$

The *Intersect* step is based on the correlation inequality below, whose proof is similar to that of Lemma 3.3.

Lemma 3.12 (Positive Correlation for Erasures - $k=0$)**.** *Let* $I_1, I_2 \subset [N]$. *Then,*

$$\mathbb{P}\Big(\bigcup_{i\in I_1} F_i \cap \bigcup_{\tilde{\imath}\in I_2} F_{\tilde{\imath}}\Big) \geq \mathbb{P}\Big(\bigcup_{i\in I_1} F_i\Big) \cdot \mathbb{P}\Big(\bigcup_{\tilde{\imath}\in I_2} F_{\tilde{\imath}}\Big). \tag{3.53}$$

In general, define $F_{i_0,\cdots,i_k}$ to be the set of all the erasure patterns such that $W_n^{(i)}$ erases for all $i \in \{i_0,\cdots,i_k\}$, i.e.,

$$F_{i_0,\cdots,i_k} = \{y \in \{0,1\}^N \mid W_n^{(i)} \text{ erases } \forall\ i \in \{i_0,\cdots,i_k\}\}, \tag{3.54}$$

and consider the set of positions SP_k containing all the subsets of k distinct elements of $\mathcal{F}^c$,

$$\mathrm{SP}_k = \{\{i_0,\cdots,i_k\} : i_m \in \mathcal{F}^c\ \forall m \in \{0,\cdots,k\}, i_m \neq i_n\ \ \forall\ m \neq n\}. \tag{3.55}$$

It it clear that

$$\mathbb{P}\Big(\bigcup_{\{i_0,\cdots,i_k\}\in \mathrm{SP}_k} F_{i_0,\cdots,i_k}\Big) = P_{\mathrm{B}}^{\mathrm{SC}}(N,R,\varepsilon,k). \tag{3.56}$$

In addition, with a small effort we generalize the result of Lemma 3.12 by following the line of thought exposed in the proof of Lemma 3.6.

Lemma 3.13 (Positive Correlation for Erasures - Any k)**.** *Let $R_1, R_2 \subset \mathrm{SP}_k$. Then,*

$$\begin{aligned}\mathbb{P}\Big(&\bigcup_{\{i_0,\cdots,i_k\}\in R_1} F_{i_0,\cdots,i_k} \cap \bigcup_{\{\tilde{i}_0,\cdots,\tilde{i}_k\}\in R_2} F_{\tilde{i}_0,\cdots,\tilde{i}_k}\Big) \\ &\geq \mathbb{P}\Big(\bigcup_{\{i_0,\cdots,i_k\}\in R_1} F_{i_0,\cdots,i_k}\Big)\cdot \mathbb{P}\Big(\bigcup_{\{\tilde{i}_0,\cdots,\tilde{i}_k\}\in R_2} F_{\tilde{i}_0,\cdots,\tilde{i}_k}\Big).\end{aligned} \tag{3.57}$$

For the *Divide* step, we need to prove that there exists $R_1 \subset \mathrm{SP}_k$, such that $\mathbb{P}(\bigcup_{\{i_0,\cdots,i_k\}\in R_1} F_{i_0,\cdots,i_k})$ is slightly less than $P_{\mathrm{B}}^{\mathrm{SC}}(N,R,\varepsilon,k)/2$. To do so, we show that, by choosing a suitably large block length, $\mathbb{P}(F_i)$ can be made as small as required. The proof of the lemma below is in Appendix 3.7.7.

Lemma 3.14 (No Big Jumps for Erasures)**.** *Let $P_{\mathrm{B}} \in (0,1)$ and $\varepsilon \in (0,1)$. Then, for $N \geq N_0(P_{\mathrm{B}},\varepsilon)$ and for any R such that*

$$\frac{1}{4} > P_{\mathrm{B}}^{\mathrm{SC}}(N,R,\varepsilon,k=0) > P_{\mathrm{B}}, \tag{3.58}$$

the probability of F_i is upper bounded by

$$\mathbb{P}(F_i) < \frac{P_{\mathrm{B}}}{8}, \qquad \forall\, i \in \mathcal{F}^c. \tag{3.59}$$

Corollary 3.5 (Existence of R_1)**.** *Let $P_{\mathrm{B}} \in (0,1)$ and $\varepsilon \in (0,1)$. Then, for N big enough and R ensuring* (3.58)*, there exists $R_1 \subset \mathrm{SP}_k$ such that*

$$\begin{aligned}\mathbb{P}\Big(\bigcup_{\{i_0,\cdots,i_k\}\in R_1} F_{i_0,\cdots,i_k}\Big) &\geq \frac{3}{8} P_{\mathrm{B}}^{\mathrm{SC}}(N,R,\varepsilon,k), \\ \mathbb{P}\Big(\bigcup_{\{i_0,\cdots,i_k\}\in R_1} F_{i_0,\cdots,i_k}\Big) &\leq \frac{1}{2} P_{\mathrm{B}}^{\mathrm{SC}}(N,R,\varepsilon,k).\end{aligned} \tag{3.60}$$

Eventually, the proof of Theorem 3.4 is obtained by using Lemma 3.13 and Corollary 3.5 and by following a procedure similar to that outlined at the end of Section 3.4.5.

3.7 Appendix

3.7.1 Proof of Lemma 3.3

Proof. Consider the Hamming space $\{0,1\}^N$. For $y, z \in \{0,1\}^N$ define the following partial order,

$$y \leq z \iff y_i \leq z_i, \qquad \forall i \in [N]. \tag{3.61}$$

Define $y \vee z$ and $y \wedge z$ as

$$\begin{aligned}(y \vee z)_i &= \begin{cases} 0 & \text{if } y_i = z_i = 0, \\ 1 & \text{else,} \end{cases} \\ (y \wedge z)_i &= \begin{cases} 1 & \text{if } y_i = z_i = 1, \\ 0 & \text{else.} \end{cases}\end{aligned} \tag{3.62}$$

Just to clarify the ideas, think of $y \in \{0,1\}^N$ as an erasure pattern, as specified at the beginning of Section 3.4.1. Since the N copies of the original BEC(ε) are independent and each of them is erased with probability ε, we consider the probability measure defined by

$$\mathbb{P}(y) = \Big(\frac{\varepsilon}{1-\varepsilon}\Big)^{w_{\mathrm{H}}(y)} (1-\varepsilon)^N, \tag{3.63}$$

where w_{H} denotes the Hamming weight.

As $w_{\mathrm{H}}(y \vee z) + w_{\mathrm{H}}(y \wedge z) = w_{\mathrm{H}}(y) + w_{\mathrm{H}}(z)$, we have

$$\mathbb{P}(y) \cdot \mathbb{P}(z) = \mathbb{P}(y \vee z) \cdot \mathbb{P}(y \wedge z). \tag{3.64}$$

For any $U_1 \subset \mathbb{F}_2^{NR}$, consider the function $f : \{0,1\}^N \to \{0,1\}$, defined as

$$f(y) = 1 - \prod_{u \in U_1} (1 - \mathbb{1}_{\{y \in E_u\}}),$$

where E_u is defined in (3.14) and $\mathbb{1}_{\{y \in E_u\}} = 1$ if and only if $y \in E_u$. Consequently, if there exists $u \in U_1$ such that $uG_y = 0$, then $f(y) = 1$; $f(y) = 0$, otherwise. Hence,

$$\mathbb{E}[f(y)] = 1 \cdot \mathbb{P}(f(y) = 1) + 0 \cdot \mathbb{P}(f(y) = 0) = \mathbb{P}\Big(\bigcup_{u \in U_1} E_u\Big).$$

If $f(y) \leq f(z)$ whenever $y \leq z$, then f is said to be monotonically increasing. If $y \leq z$, then the erasure pattern z contains all the erasures of y (and perhaps some more). Thus, if $f(y) = 1$, then $f(z) = 1$. Since f can be either 0 or 1, this is enough to show that the function is increasing.

Analogously, for any $U_2 \subset \mathbb{F}_2^{NR}$, consider the function $g : \{0,1\}^N \to \{0,1\}$ defined as

$$g(y) = 1 - \prod_{\tilde{u} \in U_2} (1 - \mathbb{1}_{\{y \in E_{\tilde{u}}\}}).$$

The function g is increasing and its expected value is given by

$$\mathbb{E}[g(y)] = \mathbb{P}\Big(\bigcup_{\tilde{u} \in U_2} E_{\tilde{u}}\Big).$$

In addition,

$$\mathbb{E}[f(y)g(y)] = \mathbb{P}\Big(\bigcup_{u \in U_1} E_u \cap \bigcup_{\tilde{u} \in U_2} E_{\tilde{u}}\Big).$$

The thesis thus follows from the version of the FKG inequality presented in Lemma 40 of [139]. □

3.7.2 Proof of Lemma 3.5

Proof. Consider the matrix $F^{\otimes n}$ defined in (1.13). Let $G = [g_1, g_2, \cdots, g_{NR}]^T$ be the generator matrix of the polar code of block length N and rate R for the BEC(ε). The matrix G is obtained by selecting the NR rows of $F^{\otimes n}$ that minimize the corresponding Bhattacharyya parameters. Then, by Lemma 3 of [53],

$$d_{\min} = \min_{1 \leq i \leq NR} w_{\mathrm{H}}(g_i),$$

where $d_{\min}$ denotes the minimum distance.

We need to show that for $n > \bar{n}(\varepsilon, P_{\mathrm{B}})$,

$$w_{\mathrm{H}}(g_i) > C(P_{\mathrm{B}}, \varepsilon), \qquad i = 1, 2, \cdots, NR, \tag{3.65}$$

where

$$C(P_{\mathrm{B}}, \varepsilon) = \frac{\ln(P_{\mathrm{B}}/8)}{\ln \varepsilon}.$$

Suppose, by contradiction, that (3.65) does not hold, i.e., there exists a row g_i such that for $n > \bar{n}(\varepsilon, P_{\mathrm{B}})$,

$$w_{\mathrm{H}}(g_i) \leq C(P_{\mathrm{B}}, \varepsilon). \tag{3.66}$$

Since G is obtained from $F^{\otimes n}$ by eliminating the rows corresponding to the frozen indices, g_i is a row of $F^{\otimes n}$, say row of index i'. Then, by Proposition 17 of [37],

$$w_{\mathrm{H}}(g_i) = 2^{w_{\mathrm{H}}(b^{(i')})} = 2^{\sum_{j=1}^{n} b_j^{(i')}},$$

where $b^{(i')} = (b_1^{(i')}, b_2^{(i')}, \cdots, b_n^{(i')})$ is the binary expansion of $i' - 1$ over n bits, $b_1^{(i')}$ being the most significant bit and $b_n^{(i')}$ the least significant bit. Consequently, (3.66) implies that

$$\sum_{j=1}^{n} b_j^{(i')} \leq \lceil \log_2 C(P_{\mathrm{B}}, \varepsilon) \rceil = c(P_{\mathrm{B}}, \varepsilon),$$

i.e., the number of 1s in the binary expansion of $i' - 1$ is upper bounded by $c(P_{\mathrm{B}}, \varepsilon)$.

Now, let us compute the Bhattacharyya parameter $Z_n^{(i')}$ of the i'-th synthetic channel. Notice that each 1 in the binary expansion of $i' - 1$ corresponds to a "+" transform and each 0 to a "−" transform. Hence,

$$Z_n^{(i')} = f_{b_1^{(i')}} \circ f_{b_2^{(i')}} \circ \cdots f_{b_n^{(i')}}(\varepsilon), \tag{3.67}$$

where $\circ$ denotes function composition and the expressions for f_0 and f_1 are deduced from (1.2),

$$f_0(x) = 2x - x^2 = 1 - (1 - x)^2, \tag{3.68}$$

$$f_1(x) = x^2. \tag{3.69}$$

Notice that f_0 and f_1 are increasing functions $\forall\, x \in [0, 1]$, and that $f_1 \circ f_0(x) \geq f_0 \circ f_1(x)\ \forall\, x \in [0, 1]$. Consequently, if we set $m = w_{\mathrm{H}}(b^{(i')})$, the minimum Bhattacharyya parameter $Z_{\min}(m)$ is obtained by applying first the function $f_1(x)$ m

times and then the function $f_0(x)$ $n-m$ times. The maximum Bhattacharyya parameter $Z_{\max}(m)$ is obtained if we apply first the function $f_0(x)$ $n-m$ times and then the function $f_1(x)$ m times. Observing also that for all $t \in \mathbb{N}$,

$$\underbrace{f_0 \circ f_0 \circ \cdots f_0(x)}_{t \text{ times}} = 1-(1-x)^{2^t}, \tag{3.70}$$

$$\underbrace{f_1 \circ f_1 \circ \cdots f_1(x)}_{t \text{ times}} = x^{2^t}, \tag{3.71}$$

we get

$$Z_{\min}(m) \leq Z_n^{(i')} \leq Z_{\max}(m), \tag{3.72}$$

with

$$Z_{\min}(m) = 1-(1-\varepsilon^{2^m})^{2^{n-m}},$$
$$Z_{\max}(m) = (1-(1-\varepsilon)^{2^{n-m}})^{2^m}.$$

Since $f_1(x) \leq f_0(x)\ \forall\ x \in [0,1]$ and $m \leq c$, we obtain that

$$Z_{\min}(m) \geq Z_{\min}(c). \tag{3.73}$$

At this point, we need to show that for k sufficiently large,

$$Z_{\min}(c) \geq Z_{\max}(c+k). \tag{3.74}$$

As $1-(1-\varepsilon)^{2^{n-c-k}} < 1$, the condition (3.74) is satisfied if

$$1-(1-\varepsilon^{2^c})^{2^{n-c}} \geq 1-(1-\varepsilon)^{2^{n-k-c}},$$

which after some simplifications leads to

$$k \geq \log_2\left(\frac{\ln(1-\varepsilon)}{\ln(1-\varepsilon^{2^c})}\right). \tag{3.75}$$

Notice that the RHS of (3.75) is an increasing function of c. As $c < \log_2(C)+1$, we deduce that the choice

$$\bar{k} = \left\lceil \log_2\left(\frac{\ln(1-\varepsilon)}{\ln(1-\varepsilon^{2C})}\right)\right\rceil = \left\lceil \log_2\left(\frac{\ln(1-\varepsilon)}{\ln(1-\varepsilon^{\frac{2\ln(P_{\mathrm{B}}/8)}{\ln\varepsilon}})}\right)\right\rceil \tag{3.76}$$

also satisfies (3.74).

An immediate consequence of inequalities (3.72), (3.73), and (3.74) is that $Z_n^{(i')} \geq Z_{\max}(c+\bar{k})$. Therefore, we can conclude that every channel of index j with $\geq c+\bar{k}$ ones in the binary expansion $b^{(j)}$ of $j-1$ has Bhattacharyya parameter $Z_n^{(j)} \leq Z_n^{(i')}$. Consequently, all these channels have not been frozen and, as $R \leq C(W) = 1-\varepsilon$,

$$\varepsilon \leq 1-R = \frac{\#\text{ frozen channels}}{\#\text{ channels}} \leq \frac{\sum_{i=0}^{c+\bar{k}-1}\binom{n}{i}}{2^n}$$
$$\leq \exp\left(\frac{-(n-2(c+\bar{k}-1))^2}{2n}\right),$$

where the last inequality is a consequence of Chernoff bound [140].

After some calculations, we conclude that for $n > \bar{n}(\varepsilon, P_{\mathrm{B}})$, where $\bar{n}(\varepsilon, P_{\mathrm{B}})$ is given by (3.26),

$$\exp\Big(\frac{-(n-2(c+\bar{k}-1))^2}{2n}\Big) < \varepsilon,$$

which is a contradiction. □

3.7.3 Proof of Lemma 3.6

Proof. As in the proof of Lemma 3.3 presented in Appendix 3.7.1, consider the Hamming space $\{0,1\}^N$ with the partial order (3.61). For $y, z \in \{0,1\}^N$ define $y \vee z$ and $y \wedge z$ as in (3.62) and take the probability measure (3.63) that satisfies (3.64). For any $P_1, P_2 \subset \mathrm{LC}_l$, pick $f : \{0,1\}^N \to \{0,1\}$ and $g : \{0,1\}^N \to \{0,1\}$, defined as

$$f(y) = 1 - \prod_{\mathrm{span}(u^{(1)},\cdots,u^{(l)})\in P_1} \big(1 - \mathbb{1}_{\{y\in E_{\mathrm{sp}(u^{(1)},\cdots,u^{(l)})}\}}\big),$$
$$g(y) = 1 - \prod_{\mathrm{span}(\tilde{u}^{(1)},\cdots,\tilde{u}^{(l)})\in P_2} \big(1 - \mathbb{1}_{\{y\in E_{\mathrm{sp}(\tilde{u}^{(1)},\cdots,\tilde{u}^{(l)})}\}}\big),$$

where $E_{\mathrm{sp}(u^{(1)},\cdots,u^{(l)})}$ is given by (3.28) and $\mathbb{1}_{\{y\in E_{\mathrm{sp}(u^{(1)},\cdots,u^{(l)})}\}} = 1$ if and only if $y \in E_{\mathrm{sp}(u^{(1)},\cdots,u^{(l)})}$. Hence,

$$\mathbb{E}[f(y)] = \mathbb{P}\Big(\bigcup_{\mathrm{span}(u^{(1)},\cdots,u^{(l)})\in P_1} E_{\mathrm{sp}(u^{(1)},\cdots,u^{(l)})}\Big),$$
$$\mathbb{E}[g(y)] = \mathbb{P}\Big(\bigcup_{\mathrm{span}(\tilde{u}^{(1)},\cdots,\tilde{u}^{(l)})\in P_2} E_{\mathrm{sp}(\tilde{u}^{(1)},\cdots,\tilde{u}^{(l)})}\Big),$$
$$\mathbb{E}[f(y)g(y)] = \mathbb{P}\Big(\bigcup_{\mathrm{span}(u^{(1)},\cdots,u^{(l)})\in P_1} E_{\mathrm{sp}(u^{(1)},\cdots,u^{(l)})} \cap \bigcup_{\mathrm{span}(\tilde{u}^{(1)},\cdots,\tilde{u}^{(l)})\in P_2} E_{\mathrm{sp}(\tilde{u}^{(1)},\cdots,\tilde{u}^{(l)})}\Big).$$

Since f and g are increasing, the thesis follows by Lemma 40 of [139]. □

3.7.4 Proof of Lemma 3.8

Proof. Assume for the moment that the output alphabet $\mathcal{Y}$ of the channel is finite and consider the binary relation $\overset{\mathcal{Y}}{\leq}$, defined for all $y_i, z_i \in \mathcal{Y}$ as

$$y_i \overset{\mathcal{Y}}{\leq} z_i \iff \frac{p(y_i|1)}{p(y_i|0)} \leq \frac{p(z_i|1)}{p(z_i|0)}. \tag{3.77}$$

The relation $\overset{\mathcal{Y}}{\leq}$ is transitive and total. As for the antisymmetry, $\overset{\mathcal{Y}}{\leq}$ satisfies the property if the following implication holds for all $y_i, z_i \in \mathcal{Y}$,

$$\frac{p(y_i|1)}{p(y_i|0)} = \frac{p(z_i|1)}{p(z_i|0)} \Longrightarrow y_i = z_i. \tag{3.78}$$

Note that, without loss of generality, we can assume that the channel output identifies with the log-likelihood ratio, see [44, Section 4.1.2]. With this assumption of

using the canonical representation of the channel, (3.78) is also fulfilled. Hence, $\overset{\mathcal{Y}}{\leq}$ is a total ordering over $\mathcal{Y}$.

Set $\mathcal{L} = \mathcal{Y}^N$ and for any $y = (y_1, \cdots, y_N)$ and $z = (z_1, \cdots, z_N)$ in $\mathcal{L}$ define the binary relation $\overset{\mathcal{L}}{\leq}$ as

$$y \overset{\mathcal{L}}{\leq} z \iff y_i \overset{\mathcal{Y}}{\leq} z_i, \qquad \forall\, i \in [N]. \tag{3.79}$$

It is easy to check that $\overset{\mathcal{L}}{\leq}$ is a partial order over the N-fold Cartesian product $\mathcal{Y}^N$.

For any $y, z \in \mathcal{L}$, denote by $y \vee z$ their unique minimal upper bound and by $y \wedge z$ their unique maximal lower bound, defined as

$$\begin{aligned}(y \vee z)_i &= \max_{\overset{\mathcal{Y}}{\leq}}(y_i, z_i), \qquad \forall\, i \in [N],\\ (y \wedge z)_i &= \min_{\overset{\mathcal{Y}}{\leq}}(y_i, z_i), \qquad \forall\, i \in [N].\end{aligned}$$

Since the distributive law holds, i.e.,

$$y \wedge (z \vee w) = (y \wedge z) \vee (y \wedge w), \qquad \forall\, y, z, w, \in \mathcal{L},$$

the set $\mathcal{L}$ with the partial ordering $\overset{\mathcal{L}}{\leq}$ is a finite distributive lattice. Observe that in the proof of Appendix 3.7.1 the finite distributive lattice $\mathcal{L}$ is replaced by the Hamming space $\{0,1\}^N$.

Let $\mu : \mathcal{L} \to \mathbb{R}^+$ be defined as

$$\mu(y) = p(y|0_{1:NR}). \tag{3.80}$$

In words, μ represents the probability of receiving the N-tuple y from the channel, given that the all-zero information vector $0_{1:NR}$ was sent. We say that such a function is log-supermodular if, for all $y, z \in \mathcal{L}$,

$$\mu(y) \cdot \mu(z) \leq \mu(y \wedge z) \cdot \mu(y \vee z). \tag{3.81}$$

An easy check shows that (3.81) is satisfied with equality with the choice (3.80). Notice that in the proof of Appendix 3.7.1 the log-supermodular function μ is replaced by the probability measure (3.63).

For any $U_1' \subset \mathbb{F}_2^{NR}$, consider the function $f : \mathcal{L} \to \{0,1\}$, defined as

$$f(y) = 1 - \prod_{u \in U_1'} (1 - \mathbb{1}_{\{y \in E_u'\}}),$$

where E_u' is given by (3.36) and $\mathbb{1}_{\{y \in E_u'\}} = 1$ if and only if $y \in E_u'$. If $f(y) \leq f(z)$ whenever $y \overset{\mathcal{L}}{\leq} z$, then f is said to be monotonically increasing. Since f can be either 0 or 1, we only need to prove the implication $f(y) = 1 \Rightarrow f(z) = 1$, whenever $y \overset{\mathcal{L}}{\leq} z$. If $f(y) = 1$, there exist $u^* \in U_1'$ such that

$$0 \leq \sum_{i \in I_{u^*}} \ln \frac{p(y_i|1)}{p(y_i|0)}.$$

As $y_i \overset{\mathcal{Y}}{\leq} z_i$ for all $i \in [N]$, by definition (3.77) we obtain

$$\sum_{i \in I_{u^*}} \ln \frac{p(y_i|1)}{p(y_i|0)} \leq \sum_{i \in I_{u^*}} \ln \frac{p(z_i|1)}{p(z_i|0)},$$

which implies that $f(z) = 1$. As a result, f is increasing.

Analogously, for any $U_2' \subset \mathbb{F}_2^{NR}$, consider the function $g : \mathcal{L} \to \{0, 1\}$ defined as

$$g(y) = 1 - \prod_{\tilde{u} \in U_2'} (1 - \mathbb{1}_{\{y \in E_{\tilde{u}}'\}}).$$

Using the same argument seen for the function f, we realize that g is an increasing function.

By the FKG inequality [141],

$$\sum_{y \in \mathcal{L}} \mu(y) f(y) \cdot \sum_{y \in \mathcal{L}} \mu(y) g(y) \leq \sum_{y \in \mathcal{L}} \mu(y) f(y) g(y) \cdot \sum_{y \in \mathcal{L}} \mu(y).$$

Observing that

$$\begin{aligned}
&\sum_{y \in \mathcal{L}} \mu(y) = 1, \\
&\sum_{y \in \mathcal{L}} \mu(y) f(y) = \mathbb{P}\Big(\bigcup_{u \in U_1'} E_u' \Big), \\
&\sum_{y \in \mathcal{L}} \mu(y) g(y) = \mathbb{P}\Big(\bigcup_{\tilde{u} \in U_2'} E_{\tilde{u}}' \Big), \\
&\sum_{y \in \mathcal{L}} \mu(y) f(y) g(y) = \mathbb{P}\Big(\bigcup_{u \in U_1'} E_u' \cap \bigcup_{\tilde{u} \in U_2'} E_{\tilde{u}}' \Big),
\end{aligned}$$

we obtain the thesis (3.38).

When the output alphabet of the channel is infinite, the proof is very similar and follows from the generalization of the FKG inequality to a finite product of totally ordered measure spaces [142].

□

3.7.5 Proof of Lemma 3.10

Proof. Following the approach of Appendix 3.7.2, suppose, by contradiction, that there is an unfrozen index i' of $F^{\otimes n}$, such that the number of 1s in the binary expansion of $i' - 1$ is upper bounded by $c(P_{\mathrm{B}}, Z)$, defined as

$$c(P_{\mathrm{B}}, Z) = \left\lceil \log_2 \frac{\ln(P_{\mathrm{B}}/8)}{\ln Z} \right\rceil.$$

The Bhattacharyya parameter $Z_n^{(i')}$ of the i'-th synthetic channel can be written as

$$Z_n^{(i')} = f_{b_1^{(i')}} \circ f_{b_2^{(i')}} \circ \cdots f_{b_n^{(i')}}(Z),$$

where the expressions for f_0 and f_1 are deduced from (1.11),

$$\sqrt{1-(1-x^2)^2} = f_0^{(L)}(x) \leq f_0(x) \leq f_0^{(U)}(x) = 1-(1-x)^2,$$
$$f_1(x) = x^2.$$

Since $f_1(x)$ and $f_0^{(L)}(x)$ are increasing and $f_0^{(L)}(x) \leq f_0(x)$, we have

$$Z_n^{(i')} \geq L_n^{(i')} \triangleq f_{b_1^{(i')}}^{(L)} \circ f_{b_2^{(i')}}^{(L)} \circ \cdots f_{b_n^{(i')}}^{(L)}(Z),$$

where, for the sake of simplicity, we have defined $f_1^{(L)}(x) = f_1(x)$. Setting $m = w_{\mathrm{H}}(b^{(i')})$ and remarking that $f_1^{(L)} \circ f_0^{(L)}(x) \geq f_0^{(L)} \circ f_1^{(L)}(x)$, a lower bound on $L_n^{(i')}$ is obtained by applying first the function $f_1^{(L)}(x)$ m times and then the function $f_0^{(L)}(x)$ $n-m$ times. Using (3.71) and observing that for all $t \in \mathbb{N}$,

$$\underbrace{f_0^{(L)} \circ f_0^{(L)} \circ \cdots f_0^{(L)}(x)}_{t \text{ times}} = \sqrt{1-(1-x^2)^{2^t}},$$

we get

$$L_n^{(i')} \geq L_{\min}(m) \triangleq \sqrt{1-(1-Z^{2^{m+1}})^{2^{n-m}}}.$$

Since $f_1^{(L)}(x) \leq f_0^{(L)}(x)$ and $m \leq c$, we obtain that

$$L_{\min}(m) \geq L_{\min}(c).$$

On the contrary, let $Z_n^{(j)}$ be the Bhattacharyya parameter of the synthetic channel of index j with $\geq c+k$ ones in the binary expansion $b^{(j)}$ of $j-1$. Since f_1 and $f_0^{(U)}$ are increasing and $f_0(x) \leq f_0^{(U)}(x)$, we have

$$Z_n^{(j)} \leq U_n^{(j)} \triangleq f_{b_1^{(j)}}^{(U)} \circ f_{b_2^{(j)}}^{(U)} \circ \cdots f_{b_n^{(j)}}^{(U)}(Z),$$

where we have defined for the sake of simplicity $f_1^{(U)}(x) = f_1(x)$. Setting $m' = w_{\mathrm{H}}(b^{(j)})$ and remarking that $f_1^{(U)} \circ f_0^{(U)}(x) \geq f_0^{(U)} \circ f_1^{(U)}(x)$, an upper bound on $U_n^{(j)}$ is obtained by applying first the function $f_0^{(U)}(x)$ $n-m'$ times and then the function $f_1^{(U)}(x)$ m' times. Using (3.70) and (3.71), we get

$$U_n^{(j)} \leq U_{\max}(m') \triangleq (1-(1-Z)^{2^{n-m'}})^{2^{m'}}.$$

Since $f_1^{(U)}(x) \leq f_0^{(U)}(x)$ and $m' \geq c+k$, we have that

$$U_{\max}(m') \leq U_{\max}(c+k).$$

At this point, we need to pick k such that the following inequality holds,

$$L_{\min}(c) \geq U_{\max}(c+k).$$

After some calculations, we obtain that

$$\bar{k} = \left\lceil \log_2 \left(\frac{\ln(1-Z)}{\ln(1-Z^{\frac{4\ln(P_{\mathrm{B}}/8)}{\ln Z}})} \right) \right\rceil$$

fulfills the requirement.

As a result, every channel of index j with $\geq c+\bar{k}$ ones in the binary expansion $b^{(j)}$ of $j-1$ cannot be frozen. By Chernoff bound [140], we get a contradiction for $n > \bar{n}(Z, C(W), P_{\mathrm{B}})$, where $\bar{n}(Z, C(W), P_{\mathrm{B}})$ is given by (3.6). □

3.7.6 Proof of Lemma 3.11

Proof. Assume at first that the output alphabet $\mathcal{Y}$ of the channel is finite and consider the finite distributive lattice $\mathcal{L} = \mathcal{Y}^N$ with the partial ordering $\overset{\mathcal{L}}{\leq}$ defined in (3.79). Let $\mu : \mathcal{L} \to \mathbb{R}^+$ be the log-supermodular function (3.80).

For any $P'_1, P'_2 \subset \mathrm{SS}_L$, consider the functions $f : \mathcal{L} \to \{0,1\}$ and $g : \mathcal{L} \to \{0,1\}$, given by

$$f(y) = 1 - \prod_{\{u^{(1)},\cdots,u^{(L)}\} \in P'_1} (1 - \mathbb{1}_{\{y \in E'_{u^{(1)},\cdots,u^{(L)}}\}}),$$
$$g(y) = 1 - \prod_{\{\tilde{u}^{(1)},\cdots,\tilde{u}^{(L)}\} \in P'_2} (1 - \mathbb{1}_{\{y \in E'_{\tilde{u}^{(1)},\cdots,\tilde{u}^{(L)}}\}}),$$

where $E'_{u^{(1)},\cdots,u^{(L)}}$ is defined in (3.42) and $\mathbb{1}_{\{y \in E'_{u^{(1)},\cdots,u^{(L)}}\}} = 1$ if and only if $y \in E'_{u^{(1)},\cdots,u^{(L)}}$. For analogous reasons to those pointed out in Appendix 3.7.4, f and g are monotonically increasing.

Noticing that

$$\sum_{y \in \mathcal{L}} \mu(y) f(y) = \mathbb{P}\Big(\bigcup_{\{u^{(1)},\cdots,u^{(L)}\} \in P'_1} E'_{u^{(1)},\cdots,u^{(L)}} \Big),$$
$$\sum_{y \in \mathcal{L}} \mu(y) g(y) = \mathbb{P}\Big(\bigcup_{\{\tilde{u}^{(1)},\cdots,\tilde{u}^{(L)}\} \in P'_2} E'_{\tilde{u}^{(1)},\cdots,\tilde{u}^{(L)}} \Big),$$
$$\sum_{y \in \mathcal{L}} \mu(y) f(y) g(y) = \mathbb{P}\Big(\bigcup_{\{u^{(1)},\cdots,u^{(L)}\} \in P'_1} E'_{u^{(1)},\cdots,u^{(L)}} \cap \bigcup_{\{\tilde{u}^{(1)},\cdots,\tilde{u}^{(L)}\} \in P'_2} E'_{\tilde{u}^{(1)},\cdots,\tilde{u}^{(L)}} \Big),$$

the thesis follows from the FKG inequality [141]. To handle the case of an infinite output alphabet, it is enough to apply the generalization of the FKG inequality in [142].

□

3.7.7 Proof of Lemma 3.14

Proof. Suppose that the thesis does not hold, i.e.,

$$\max_{i \in \mathcal{F}^c} \mathbb{P}(F_i) = \max_{i \in \mathcal{F}^c} Z_i = \alpha \geq \frac{P_{\mathrm{B}}}{8}.$$

Consider $a, b \in (0,1)$ that satisfy

$$\sqrt{a} \leq 1 - \sqrt{1-b}. \tag{3.82}$$

Then, for any $\varepsilon \in (0,1)$ and for N sufficiently large, by Corollary 6 of [128] the number of channels $N_c(a,b,N,\varepsilon)$ whose Bhattacharyya parameter is contained in the interval $[a,b]$ is lower bounded by $N^{1+\lambda^{(l)}_{\mathrm{BEC}}}$, where $\lambda^{(l)}_{\mathrm{BEC}} \geq -0.279$. Since the choice $b = \alpha$ and $a = (\alpha/2)^2$ satisfies (3.82), we obtain

$$N_c\left(\left(\frac{\alpha}{2}\right)^2, \alpha, N, \varepsilon\right) \geq A = \lfloor N^{1+\lambda^{(l)}_{\mathrm{BEC}}} \rfloor. \tag{3.83}$$

Let B_i be the erasure indicator of the i-th synthetic channel of Bhattacharyya parameter $Z_n^{(i)}$. Then, $B_i \in \{0,1\}$ is a binary random variable such that $\mathbb{P}(B_i = 1) = Z_n^{(i)}$. Denote by $\rho_{i,j}$ the correlation coefficient between the erasure indicators of the i-th and the j-th channel. This correlation coefficient can be expressed as

$$\rho_{i,j} = \frac{\mathbb{E}(B_i B_j) - \mathbb{E}(B_i)\mathbb{E}(B_j)}{\mathrm{var}(B_i)\mathrm{var}(B_j)}.$$

By Corollary 2 of [143], we have that

$$\sum_{i,j \in \{1,\cdots,N\}} \rho_{i,j} \leq N^{3-\log_2(3)}. \tag{3.84}$$

Let $\mathcal{A}_{\max}$ be the set of indices of the unfrozen channels with the highest Bhattacharyya parameters such that $|\mathcal{A}_{\max}| = A$. Notice that the Bhattacharyya parameters of these channels are contained in the interval $[(\alpha/2)^2, \alpha]$ by (3.83). Denote by R_A the associated $A \times A$ matrix of the correlation coefficients. We are going to show that for any $M \in \mathbb{N}$, there exists $S_M^* \subset \mathcal{A}_{\max}$, with $|S_M^*| = M$, such that

$$\max_{\substack{i,j \in S_M^* \\ i \neq j}} \rho_{i,j} < \binom{M}{2} \frac{N^{3-\log_2(3)}}{A^2}. \tag{3.85}$$

Since $3 - \log_2(3) - 2(1 + \lambda_{\mathrm{BEC}}^{(l)}) < 0$, the previous relation implies that, if we fix M and we choose N suitably large, then the correlation coefficients of the channels with indices in S_M^* can be made arbitrarily small.

To prove (3.85), first observe that (3.84) clearly implies that $\sum_{i,j \in \mathcal{A}_{\max}} \rho_{i,j} \leq N^{3-\log_2(3)}$. Hence, the average of all the elements of the matrix R_A is upper bounded by $N^{3-\log_2(3)}/A^2$. As R_A is symmetric and its principal diagonal is made up by 1s, the average of the strictly upper triangular part of R_A, i.e., the average of the $\binom{A}{2}$ elements of R_A that are above the principal diagonal, is also upper bounded by $N^{3-\log_2(3)}/A^2$. In formulae,

$$\frac{1}{\binom{A}{2}} \sum_{\substack{i,j \in \mathcal{A}_{\max} \\ i<j}} \rho_{i,j} \leq \frac{N^{3-\log_2(3)}}{A^2}.$$

Let S_M be a subset of $\mathcal{A}_{\max}$ with cardinality $|S_M| = M$. We can associate with S_M the $\binom{M}{2}$ elements of the strictly upper triangular part of R_A, which represent the correlation coefficients of the channels whose indices are in S_M. By symmetry, when we consider all the subsets of cardinality M of $\mathcal{A}_{\max}$, we count each element of the strictly upper triangular part of R_A the same number of times, i.e., $\binom{A-2}{M-2}$. As a result, by noticing that there are $\binom{A}{M}$ distinct subsets of cardinality M of $\mathcal{A}_{\max}$, we have

$$\frac{1}{\binom{A}{M}} \sum_{S_M \subset \mathcal{A}_{\max}} \frac{1}{\binom{M}{2}} \sum_{\substack{i,j \in S_M \\ i<j}} \rho_{i,j} \leq \frac{N^{3-\log_2(3)}}{A^2}.$$

Consequently, there exists $S_M^* \subset \mathcal{A}_{\max}$, such that

$$\frac{1}{\binom{M}{2}} \sum_{\substack{i,j \in S_M^* \\ i<j}} \rho_{i,j} \leq \frac{N^{3-\log_2(3)}}{A^2},$$

which implies (3.85).

With the choice $M = \lceil 128/P_{\mathrm{B}}^2 \rceil$, it is easy to see that there exists $S^* \subset S_M^*$ that satisfies

$$\frac{1}{2} + \alpha \geq \sum_{i \in S^*} Z_n^{(i)} \geq \frac{1}{2}. \tag{3.86}$$

Indeed, $\sum_{i \in S_M^*} Z_n^{(i)} \geq M(\alpha/2)^2 \geq 1/2$ and $\max_{i \in S_M^*} Z_n^{(i)} \leq \alpha$.

An application of Bonferroni's inequality (see [144, Section 4.7]) yields

$$P_{\mathrm{B}}^{\mathrm{SC}}(N, R, \varepsilon, k = 0) \geq \mathbb{P}\Big(\bigcup_{i \in S^*} F_i\Big) \geq \sum_{i \in S^*} \mathbb{P}(F_i) - \frac{1}{2} \sum_{\substack{i,j \in S_M^* \\ i \neq j}} \mathbb{P}(F_i \cap F_j). \tag{3.87}$$

The term $\mathbb{P}(F_i \cap F_j)$ can be upper bounded as

$$\begin{aligned} \mathbb{P}(F_i \cap F_j) &= Z_n^{(i)} Z_n^{(j)} + \rho_{i,j} \sqrt{Z_n^{(i)} Z_n^{(j)} (1 - Z_n^{(i)})(1 - Z_n^{(j)})} \\ &\overset{\text{(a)}}{\leq} Z_n^{(i)} Z_n^{(j)} + \binom{M}{2} \frac{N^{3 - \log_2(3)}}{A^2} \\ &\overset{\text{(b)}}{\leq} Z_n^{(i)} Z_n^{(j)} + \frac{1}{8\binom{M}{2}}, \end{aligned} \tag{3.88}$$

where the inequality (a) comes from (3.85) and the fact that $Z_n^{(i)} \in [0, 1]$; and the inequality (b) is easily obtained by picking N large enough.

Using (3.87) and (3.88), we have

$$P_{\mathrm{B}}^{\mathrm{SC}}(N, R, \varepsilon, k = 0) \geq \sum_{i \in S^*} Z_n^{(i)} - \frac{1}{2}\Big(\sum_{i \in S^*} Z_n^{(i)}\Big)^2 - \frac{1}{8}. \tag{3.89}$$

Note that

$$\alpha \leq P_{\mathrm{B}}^{\mathrm{SC}}(N, R, \varepsilon, k) < \frac{1}{4},$$

where the last inequality comes from the hypothesis of the Lemma. Hence, by using (3.86), we deduce that

$$\sum_{i \in S^*} Z_n^{(i)} < \frac{3}{4} < 1.$$

Since the function $h(x) = x - x^2/2$ is increasing in $[0, 1]$ and $1/2 \leq \sum_{i \in S^*} Z_n^{(i)} < 1$, we can conclude that

$$\sum_{i \in S^*} Z_n^{(i)} - \frac{1}{2}\Big(\sum_{i \in S^*} Z_n^{(i)}\Big)^2 - \frac{1}{8} \geq \frac{1}{4}, \tag{3.90}$$

which is a contradiction and gives us the thesis.

□

How to Achieve Marton's Region for Broadcast Channels

4

Usa meno virgolette possibili: non è "fine".

Limit the use of inverted commas: it is not "elegant".

The second topic of this thesis concerns the design of practical coding techniques for non-standard communication scenarios. In this chapter[1], we focus on polar codes for the broadcast channel.

In Section 4.1, we introduce two fundamental transmission strategies for the broadcast setting, i.e., superposition coding and binning. Then, in Section 4.2, we state our main result: we present polar coding schemes that achieve with low-complexity the best inner bound known to date, i.e., Marton's region. To achieve this goal, we build on the polar coding solution proposed in [100], and we develop a chaining construction that enables us to remove the degradation assumptions of [100]. In Section 4.3, we characterize explicitly the rate regions obtained with the classic information-theoretic schemes, i.e., the superposition region, the binning region, and Marton's region. We also review the rate regions achievable by the polar constructions of [100] and provide an explicit example of a case in which the extra degradation assumptions do not enable us to obtain the full information-theoretic region. In Section 4.4, we describe two crucial polar "primitives": polar coding schemes for lossless compression with and without side information, and for transmission over binary memoryless channels, either symmetric or asymmetric. In Sections 4.5 and 4.6, we present our new polar coding schemes to achieve the entire superposition and binning regions, respectively. Finally, in Section 4.7, we combine these two constructions in order to achieve the whole Marton's region, both with private and common messages.

[1] The material of this chapter is based on joint work with S. H. Hassani, I. Sason, and R. Urbanke [145, 146].

4.1 Related Work

Goela, Abbe, and Gastpar recently introduced polar coding schemes for the m-user deterministic broadcast channel and for the noisy discrete memoryless broadcast channel (DM-BC) [100]. For the second scenario, they considered two fundamental transmission strategies: *superposition coding*, in the version proposed by Bergmans [147], and *binning* [148]. In order to guarantee a proper alignment of the polar indices, in both the superposition and binning schemes, their solution involves some degradation constraints that are assumed to hold between the auxiliary random variables and the channel outputs.

Originally, two superposition coding schemes were proposed by Bergmans [147] and by Cover [149], and they both achieve the capacity region of the degraded broadcast channel. However, it has recently been proved that under MAP decoding, Cover's strategy always achieves a rate region at least as large as Bergmans', and this dominance can sometimes be strict [150].

The original work by Marton on binning [148] covers the case with only private messages, and the introduction of common information is due to Gelfand and Pinsker [151]. Hence, we will refer to this region as the Marton-Gelfand-Pinsker (MGP) region. This rate region is tight for all classes of DM-BCs with known capacity region, and it forms the best inner bound known to date for a two-user DM-BC [152, 153]. Note that the MGP region also includes Cover's superposition region.

4.2 Main Result

The contribution of this chapter consists in showing **how to achieve Marton's region** with both common and private messages, i.e., the MGP region, with a **practical low-complexity scheme**.

In order to extend the polar schemes of [100] that will be denoted as AGG constructions, the crucial point consists in removing the degradation conditions on auxiliary random variables and channel outputs. Note that, in general, such kind of extra conditions make the achievable rate region strictly smaller, see [154]. The ideas that make it possible to lift the constraints come from recent progress in constructing *universal* polar codes that are capable of achieving the compound capacity of the whole class of BMS channels [48, 49]. A similar technique has also been used to achieve strong security on degraded wiretap channels [110].

In short, first we describe polar codes for the superposition and binning strategies. Then, by combining these two techniques, we achieve Marton's rate region with private messages only. Finally, by describing how to transmit common information, we achieve the whole MGP region. The proposed schemes possess the standard properties of polar codes: the encoding and decoding complexity is $\Theta(N \log_2 N)$, where N is the block length, and the block error probability scales as $O(2^{-N^\beta})$, for any $\beta \in (0, 1/2)$.

4.3 Achievable Rate Regions

In this section, we review the regions achieved by the classic information-theoretic schemes and by the recent AGG polar coding constructions. We also provide an

example of a simple case in which the AGG superposition region is strictly smaller than the one obtainable by information theoretic arguments. This motivates us to build low-complexity codes capable of transmitting reliably at any rate pair inside the larger information-theoretic regions. To simplify such a task, we show that, in order to achieve the full information-theoretic regions, it suffices to achieve few specific rate pairs.

4.3.1 Information-Theoretic Schemes

Let us start by considering the rate region that is achievable by Bergmans' superposition scheme and that provides the capacity region of degraded DM-BCs (see Theorem 5.1 of [152]).

Theorem 4.1 (Superposition Region)**.** *Consider the transmission over a two-user DM-BC $p_{Y_1,Y_2|X}$, where X denotes the input to the channel, and Y_1, Y_2 denote the outputs at the first and second receiver, respectively. Let V be an auxiliary random variable. Then, for any joint distribution $p_{V,X}$ such that $V - X - (Y_1, Y_2)$ forms a Markov chain, a rate pair (R_1, R_2) is achievable if*

$$\begin{aligned} R_1 &< I(X;Y_1 \mid V), \\ R_2 &< I(V;Y_2), \\ R_1 + R_2 &< I(X;Y_1). \end{aligned} \tag{4.1}$$

Note that the above only describes a subset of the region actually achievable by superposition coding. We get a second subset by swapping the roles of the two users, i.e., by swapping the indices 1 and 2. The actual achievable region is obtained by the convex hull of the closure of the union of these two subsets.

The rate region, which is achievable by the binning strategy, is described by the following result (see Theorem 8.3 of [152]).

Theorem 4.2 (Binning Region)**.** *Consider the transmission over a two-user DM-BC $p_{Y_1,Y_2|X}$, where X denotes the input to the channel, and Y_1, Y_2 denote the outputs at the first and second receiver, respectively. Let V_1 and V_2 denote auxiliary random variables. Then, for any joint distribution p_{V_1,V_2} and for any deterministic function ϕ such that $X = \phi(V_1, V_2)$, a rate pair (R_1, R_2) is achievable if*

$$\begin{aligned} R_1 &< I(V_1;Y_1), \\ R_2 &< I(V_2;Y_2), \\ R_1 + R_2 &< I(V_1;Y_1) + I(V_2;Y_2) - I(V_1;V_2). \end{aligned} \tag{4.2}$$

Note that the achievable rate region does not become larger by considering general distributions $p_{X|V_1,V_2}$, i.e., there is no loss of generality in restricting X to be a deterministic function of (V_1, V_2), as shown in Remark 8.4 of [152]. Furthermore, for deterministic DM-BCs, the choice $V_1 = Y_1$ and $V_2 = Y_2$ in (4.2) provides their capacity region, as shown in Example 7.1 of [153].

The rate region in (4.2) can be enlarged by combining binning with superposition coding. This leads to Marton's region for a two-user DM-BC where only private messages are available, see Theorem 2 of [148] and Proposition 8.1 of [152].

Theorem 4.3 (Marton's Region)**.** *Consider the transmission over a two-user DM-BC $p_{Y_1,Y_2|X}$, where X denotes the input to the channel, and Y_1, Y_2 denote the outputs at the first and second receiver, respectively. Let V, V_1, and V_2 denote auxiliary random variables. Then, for any joint distribution p_{V,V_1,V_2} and for any deterministic function ϕ such that $X = \phi(V, V_1, V_2)$, a rate pair (R_1, R_2) is achievable if*

$$\begin{aligned} R_1 &< I(V, V_1; Y_1), \\ R_2 &< I(V, V_2; Y_2), \\ R_1 + R_2 &< I(V, V_1; Y_1) + I(V_2; Y_2 \mid V) - I(V_1; V_2 \mid V), \\ R_1 + R_2 &< I(V, V_2; Y_2) + I(V_1; Y_1 \mid V) - I(V_1; V_2 \mid V). \end{aligned} \tag{4.3}$$

Note that the binning region (4.2) is a special case of Marton's region (4.3) where the random variable V is set to be a constant. As for the binning region in Theorem 4.2, there is no loss of generality in restricting X to be a deterministic function of (V, V_1, V_2).

In a more general set-up, the users can transmit also common information. The generalization of Theorem 4.3 to the case with a common message yields the MGP region. We denote by R_0 the rate associated with the common message, and R_1, R_2 continue to indicate the private rates of the first and the second user, respectively. Then, under the hypotheses of Theorem 4.3, a rate triple (R_0, R_1, R_2) is achievable if

$$\begin{aligned} R_0 &< \min\{I(V; Y_1), I(V; Y_2)\}, \\ R_0 + R_1 &< I(V, V_1; Y_1), \\ R_0 + R_2 &< I(V, V_2; Y_2), \\ R_0 + R_1 + R_2 &< I(V, V_1; Y_1) + I(V_2; Y_2 \mid V) - I(V_1; V_2 \mid V), \\ R_0 + R_1 + R_2 &< I(V, V_2; Y_2) + I(V_1; Y_1 \mid V) - I(V_1; V_2 \mid V). \end{aligned} \tag{4.4}$$

An equivalent form of this region was derived in [155–157] (see also Theorem 8.4 and Remark 8.6 in [152]). Note that the MGP region (4.4) is specialized into Marton's region (4.3) when $R_0 = 0$, i.e., if only private messages are considered. The evaluation of Marton's region in (4.3) and of the MGP region in (4.4) for DM-BCs was studied in [158–160]. Furthermore, the optimality of these regions was proved in [161, 162] for some interesting models of broadcast channels.

4.3.2 Existing Polar Constructions

Let us now compare the results of Theorems 4.1 and 4.2 with the superposition and binning regions that are achievable by the polarization-based constructions in [100]. We write $p \succ q$ to indicate that the channel q is stochastically degraded with respect to the channel p.

Theorem 4.4 (AGG Superposition Region)**.** *Consider the transmission over a two-user DM-BC $p_{Y_1,Y_2|X}$ with a binary-input alphabet, where X denotes the input to the channel, and Y_1, Y_2 denote the outputs at the first and second receiver, respectively. Let V be an auxiliary binary random variable and assume that $p_{Y_1|V} \succ p_{Y_2|V}$. Then, for any joint distribution $p_{V,X}$ such that $V - X - (Y_1, Y_2)$ forms a Markov chain and*

for any rate pair (R_1, R_2) *such that*

$$\begin{aligned} R_1 &< I(X; Y_1 \mid V), \\ R_2 &< I(V; Y_2), \end{aligned} \tag{4.5}$$

there exists a sequence of polar codes with an increasing block length N *that achieves this rate pair with encoding and decoding complexity* $\Theta(N \log_2 N)$*, and with a block error probability that decays like* $O(2^{-N^\beta})$ *for any* $\beta \in (0, 1/2)$*.*

Theorem 4.5 (AGG Binning Region)**.** *Consider the transmission over a two-user DM-BC* $p_{Y_1,Y_2|X}$*, where* X *denotes the input to the channel, and* Y_1*,* Y_2 *denote the outputs at the first and second receiver, respectively. Let* V_1 *and* V_2 *denote auxiliary binary random variables and assume that* $p_{Y_2|V_2} \succ p_{V_1|V_2}$*. Then, for any joint distribution* p_{V_1,V_2}*, for any deterministic function* ϕ *such that* $X = \phi(V_1, V_2)$*, and for any rate pair* (R_1, R_2) *such that*

$$\begin{aligned} R_1 &< I(V_1; Y_1), \\ R_2 &< I(V_2; Y_2) - I(V_1; V_2), \end{aligned} \tag{4.6}$$

there exists a sequence of polar codes with an increasing block length N *that achieves this rate pair with encoding and decoding complexity* $\Theta(N \log_2 N)$*, and with a block error probability that decays like* $O(2^{-N^\beta})$ *for any* $\beta \in (0, 1/2)$*.*

The rate regions (4.5) and (4.6) describe a subset of the regions actually achievable with polar codes by superposition coding and binning, respectively. We get a second subset by swapping the roles of the two users. However, in some cases it is not possible to achieve this second subset, as, by swapping the indices 1 and 2, we might not be able to fulfill the required degradation assumptions.

4.3.3 Comparison of Superposition Regions

For motivation, before proceeding with the new code constructions and proofs, let us consider a specific transmission scenario and compare the information-theoretic superposition region (4.1) and the AGG superposition region (4.5) where the latter requires the degradation assumption $p_{Y_1|V} \succ p_{Y_2|V}$.

In the following, let the channel between X and Y_1 be a binary symmetric channel with crossover probability p, namely, a BSC(p), and the channel between X and Y_2 be a binary erasure channel with erasure probability ε, namely, a BEC(ε). Let us recall a few known results for this specific model (see Example 5.4 of [152]).

1. For any choice of the parameters $p \in (0, 1/2)$ and $\varepsilon \in (0, 1)$, the capacity region of this DM-BC is achieved using superposition coding.

2. For $0 < \varepsilon < 2p$, Y_1 is a stochastically degraded version of Y_2.

3. For $4p(1-p) < \varepsilon \leq h_2(p)$, Y_2 is more capable than Y_1, i.e. $I(X; Y_2) \geq I(X; Y_1)$ for all distributions p_X, where $h_2(p) = -p \log_2 p - (1-p) \log_2 (1-p)$ denotes the binary entropy function.

Let $\mathcal{V}$ and $\mathcal{X}$ denote the alphabets of the auxiliary random variable V and of the input X, respectively. Then, if the DM-BC is stochastically degraded or more capable, the auxiliary random variables satisfy the cardinality bound $|\mathcal{V}| \leq |\mathcal{X}|$ [163]. Consequently, for such a set of parameters, we can restrict our analysis to binary auxiliary random variables without any loss of generality. Furthermore, from Lemma 7 of [164], we can assume that the channel from V to X is a BSC and that the binary random variable X is symmetric.

First, pick $p = 0.11$ and $\varepsilon = 0.2$. In this case, the DM-BC is stochastically degraded and, as can be seen in Figure 4.1a, the two regions (4.1) and (4.5) coincide despite of the presence of the extra degradation assumption. In addition, these two regions are non-trivial, in the sense that they improve upon the simple time-sharing scheme in which one user remains silent and the other employs a point-to-point capacity-achieving code. Then, pick $p = 0.11$ and $\varepsilon = 0.4$. In the latter case, the DM-BC is more capable and, as can be seen in Figure 4.1b, the information-theoretic region (4.1) strictly improves upon the AGG region (4.5) that coincides with a trivial time-sharing.

The example above shows that the degradation conditions needed by the existing polar constructions strictly shrink the achievable rate region. This motivates us to achieve the information-theoretic regions described in Section 4.3.1 by a means of a practical low-complexity coding scheme.

4.3.4 Equivalent Description of Achievable Regions

When describing our new polar coding schemes, we will show how to achieve certain rate pairs. The following propositions state that the achievability of these rate pairs is equivalent to the achievability of the whole rate regions described in Theorems 4.1–4.3.

Proposition 4.1 (Equivalent Superposition Region)**.** *In order to show the achievability of all points in the region* (4.1)*, it suffices to describe a sequence of codes with an increasing block length N that achieves each of the rate pairs*

- $(R_1, R_2) = (I(X;Y_1 \mid V), \min(I(V;Y_1), I(V;Y_2)))$,
- $(R_1, R_2) = (I(X;Y_1) - I(V;Y_2), I(V;Y_2))$*, provided that* $I(V;Y_1) < I(V;Y_2) < I(X;Y_1)$,

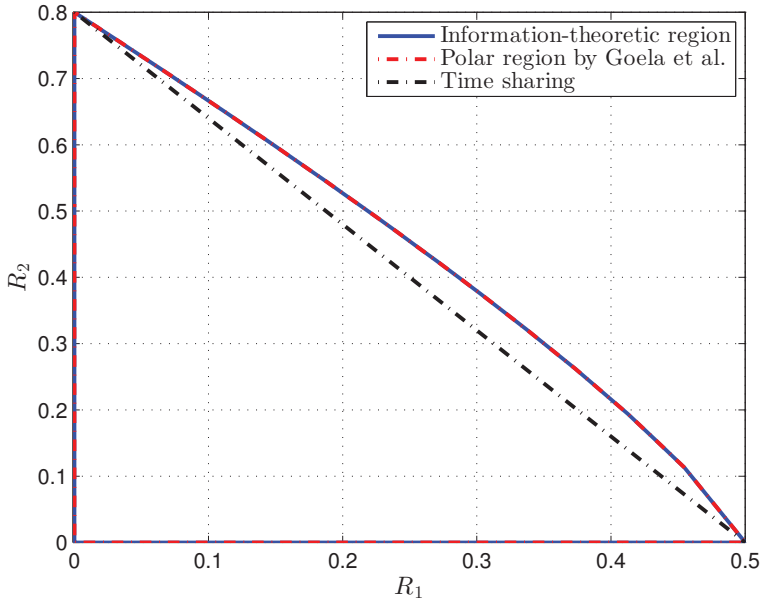
with a block error probability that decays to zero as $N \to \infty$.

Proof. We first assume that $I(V;Y_2) \leq I(V;Y_1)$. Since $V - X - Y_1$ forms a Markov chain, by the chain rule, the first two inequalities in (4.1) imply that

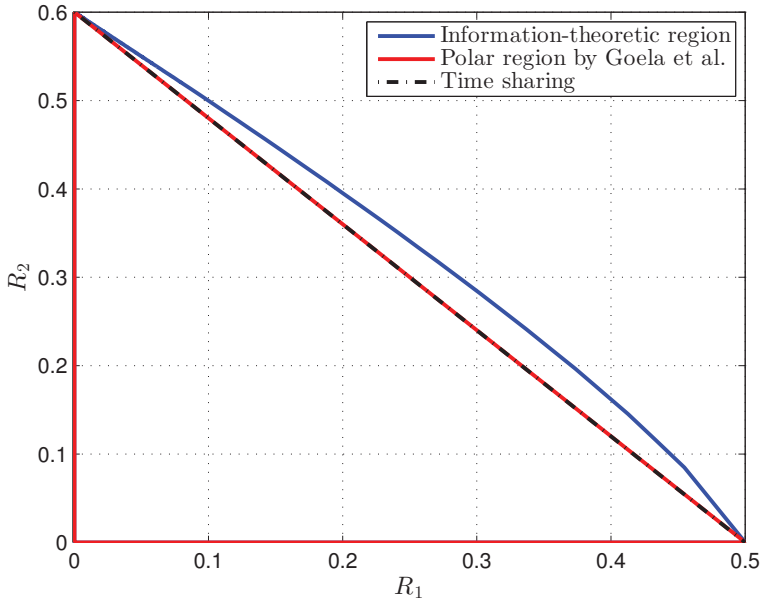
$$\begin{aligned} R_1 + R_2 &< I(X;Y_1 \mid V) + I(V;Y_2) \\ &\leq I(X;Y_1 \mid V) + I(V;Y_1) \\ &= I(V,X;Y_1) \\ &= I(X;Y_1). \end{aligned}$$

Hence, the region (4.1) is a rectangle and it suffices to achieve the corner point $(I(X;Y_1 \mid V), I(V;Y_2))$.

Now, suppose that $I(V;Y_1) < I(V;Y_2)$. Let us separate this case into the following two sub-cases:



(a) $\varepsilon = 0.2$



(b) $\varepsilon = 0.4$

Figure 4.1 – Comparison of superposition regions when the channel from X to Y_1 is a BSC(0.11) and the channel from X to Y_2 is a BEC(ε). When $\varepsilon = 0.2$, the information-theoretic region (in blue) coincides with the AGG region (in red) and they are both strictly larger than the time-sharing line (in black). When $\varepsilon = 0.4$, the information-theoretic region is strictly larger than the AGG region that reduces to the time-sharing line.

1. If $I(X;Y_1) > I(V;Y_2)$, the region (4.1) is a pentagon with the corner points

$$(I(X;Y_1) - I(V;Y_2), I(V;Y_2)), (I(X;Y_1 \mid V), I(V;Y_1)).$$

The reason for the first corner point is that $I(V;Y_1 \mid X) = 0$. Thus, if $R_2 = I(V;Y_2)$, the satisfiability of the equality $R_1 + R_2 = I(X;Y_1)$ yields that

$$\begin{aligned} R_1 &= I(X;Y_1) - I(V;Y_2) \\ &= I(V,X;Y_1) - I(V;Y_2) \\ &< I(V,X;Y_1) - I(V;Y_1) \\ &= I(X;Y_1 \mid V). \end{aligned}$$

The reason for the second corner point is that $R_1 = I(X;Y_1 \mid V)$, $R_2 = I(V;Y_1) < I(V;Y_2)$, and

$$\begin{aligned} R_1 + R_2 &= I(V,X;Y_1) \\ &= I(V;Y_1 \mid X) + I(X;Y_1) \\ &= I(X;Y_1). \end{aligned}$$

2. Otherwise, if $I(X;Y_1) \leq I(V;Y_2)$, the region (4.1) is a right trapezoid with corner points $(I(X;Y_1 \mid V), I(V;Y_1))$ and $(0, I(X;Y_1))$. As $V - X - Y_2$ forms a Markov chain, by the data processing theorem and the last condition, it follows that $I(X;Y_1) \leq I(V;Y_2) \leq I(X;Y_2)$. Hence, the second corner point $(0, I(X;Y_1))$ is dominated by the point achievable when the first user is kept silent and the second user adopts a reliable point-to-point code with rate close to $I(X;Y_2)$.

□

Proposition 4.2 (Equivalent Binning Region)**.** *In order to show the achievability of all points in the region* (4.2)*, it suffices to describe a sequence of codes with an increasing block length N that achieves the rate pair*

$$(R_1, R_2) = (I(V_1;Y_1), I(V_2;Y_2) - I(V_1;V_2)),$$

assuming that $I(V_1;V_2) \leq I(V_2;Y_2)$, with a block error probability that decays to zero as $N \to \infty$.

Proof. Assume that

$$I(V_1;V_2) \leq \min(I(V_1;Y_1), I(V_2;Y_2)).$$

Then, the region (4.2) is a pentagon with corner points

$$(I(V_1;Y_1), I(V_2;Y_2) - I(V_1;V_2)), (I(V_1;Y_1) - I(V_1;V_2), I(V_2;Y_2)).$$

Since the region (4.2) and the above condition are not affected by swapping the indices 1 and 2, it suffices to achieve the first corner point. In order to obtain the other corner point, we simply exchange the roles of the two users.

Next, suppose that $I(V_2;Y_2) \le I(V_1;V_2) < I(V_1;Y_1)$. Then, the region (4.2) is a right trapezoid with corner points

$$(I(V_1;Y_1) - I(V_1;V_2), I(V_2;Y_2)), (I(V_1;Y_1) + I(V_2;Y_2) - I(V_1;V_2), 0).$$

Note that

$$I(V_1;Y_1) + I(V_2;Y_2) - I(V_1;V_2) \le I(V_1;Y_1) \le I(X;Y_1),$$

where the last inequality follows from the data processing theorem for the Markov chain $V_1 - X - Y_1$. Hence, the rate pair $(I(V_1;Y_1) + I(V_2;Y_2) - I(V_1;V_2), 0)$ is dominated by the point $(I(X;Y_1), 0)$ that is achievable when the first user adopts a reliable point-to-point code with rate close to $I(X;Y_1)$ and the second user is kept silent.

The case where $I(V_1;Y_1) \le I(V_1;V_2) < I(V_2;Y_2)$ is solved by swapping the indices of the two users, and by referring to the previous case.

Finally, assume that $I(V_1;V_2) \ge \max(I(V_1;Y_1), I(V_2;Y_2))$. Then, the region (4.2) is a triangle with corner points that are achievable when one user is kept silent, whereas the other user adopts a reliable point-to-point code. □

Proposition 4.3 (Equivalent Marton's Region)**.** *In order to show the achievability of all points in the region* (4.3)*, it suffices to describe a sequence of codes with an increasing block length N that achieves each of the rate pairs*

$$\begin{aligned}(R_1, R_2) &= (I(V, V_1; Y_1), I(V_2; Y_2 \mid V) - I(V_1; V_2 \mid V)),\\ (R_1, R_2) &= (I(V, V_1; Y_1) - I(V_1; V_2 \mid V) - I(V; Y_2), I(V, V_2; Y_2)),\end{aligned} \tag{4.7}$$

assuming that $I(V;Y_1) \le I(V;Y_2)$, with a block error probability that decays to zero as $N \to \infty$.

Proof. Since the region (4.3) is not affected by swapping the indices 1 and 2, we can assume without loss of generality that $I(V;Y_1) \le I(V;Y_2)$. Then,

$$\begin{aligned}I(V, V_1; Y_1) + I(V_2; Y_2 \mid V) &= I(V;Y_1) + I(V_1; Y_1|V) + I(V_2; Y_2 \mid V)\\ &\le I(V;Y_2) + I(V_1; Y_1|V) + I(V_2; Y_2 \mid V)\\ &= I(V, V_2; Y_2) + I(V_1; Y_1 \mid V),\end{aligned}$$

which means that the fourth inequality in (4.3) does not restrict the rate region under the above assumption.

Now, we can follow the same procedure outlined in the proof of Propositions 4.1 and 4.2. Suppose that

$$\begin{aligned}&I(V_2; Y_2 \mid V) - I(V_1; V_2 \mid V) > 0,\\ &I(V, V_1; Y_1) - I(V_1; V_2 \mid V) - I(V; Y_2) > 0.\end{aligned} \tag{4.8}$$

Then, the rate region (4.3) is a pentagon with the corner points in (4.7).

If one of the inequalities in (4.8) is satisfied and the other is violated, then the region (4.3) is a right trapezoid whose corner points can be obtained as follows: the first corner point is given by (4.7); and the second corner point is obtained when one user remains silent and the other uses a point-to-point reliable code.

If both inequalities in (4.8) are violated, then the region (4.3) is a triangle with corner points that are achievable with reliable point-to-point codes. □

4.4 Polar Coding Primitives

The AGG constructions, as well as our extensions, are based on two polar coding "primitives". Therefore, before discussing the broadcast setting, let us review these basic scenarios.

The first such primitive is the lossless compression, with or without side information. In the polar setting, this problem was first discussed in [42, 53]. In Section 4.4.1, we consider the point of view of source polarization, described in [91, 92].

The second such primitive is the transmission of polar codes over a binary memoryless channel that can be either symmetric or asymmetric. We will consider this problem in detail in the next chapter, where we will present three different coding paradigms to solve it. For the moment, let us just point out that the basic issue consists in the fact that linear codes impose a uniform input distribution, whereas the capacity-achieving input distribution is in general not uniform when the channel is asymmetric. A solution that makes use of the concatenation of two polar codes was proposed in [165]. However, a more direct polar scheme is implicitly considered in [100], and it is independently and explicitly presented in [121]. We will briefly review this last approach in Section 4.4.2.

4.4.1 Lossless Compression

Problem Statement. Consider a binary random variable $X \sim p_X$. Then, given the random vector $X_{1:N} = (X_1, \cdots, X_N)$ consisting of N i.i.d. copies of X, the aim is to compress $X_{1:N}$ in a lossless fashion into a binary codeword of size roughly $NH(X)$, which is the entropy of $X_{1:N}$.

Design of the Scheme. Let $U_{1:N} = (U_1, \cdots, U_N)$ be defined as

$$U_{1:N} = X_{1:N} G_N, \tag{4.9}$$

where G_N is defined in (1.13). Then, $U_{1:N}$ is a random vector whose components are polarized in the sense that, with high probability, either U_i is approximately uniform and independent of $U_{1:i-1}$, or U_i is approximately a deterministic function of $U_{1:i-1}$. Formally, for $\beta \in (0, 1/2)$, let $\delta_N = 2^{-N^\beta}$ and set

$$\begin{aligned} \mathcal{H}_X &= \{i \in [N] \colon Z(U_i \mid U_{1:i-1}) \geq 1 - \delta_N\}, \\ \mathcal{L}_X &= \{i \in [N] \colon Z(U_i \mid U_{1:i-1}) \leq \delta_N\}, \end{aligned} \tag{4.10}$$

where Z denotes the Bhattacharyya parameter defined in (1.16). Hence, for $i \in \mathcal{H}_X$, the bit U_i is approximately uniformly distributed and independent of the past $U_{1:i-1}$; and, for $i \in \mathcal{L}_X$, the bit U_i is approximately a deterministic function of $U_{1:i-1}$. Furthermore,

$$\begin{aligned} \lim_{N\to\infty} \frac{1}{N} |\mathcal{H}_X| &= H(X), \\ \lim_{N\to\infty} \frac{1}{N} |\mathcal{L}_X| &= 1 - H(X). \end{aligned} \tag{4.11}$$

For a graphical representation of this setting, see Figure 4.2.

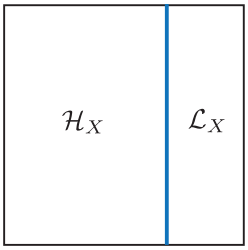


Figure 4.2 – A simple graphical representation of the sets $\mathcal{H}_X$ and $\mathcal{L}_X$ for the lossless compression scheme. The whole square represents $[N]$. The sets $\mathcal{H}_X$ and $\mathcal{L}_X$ almost form a partition of $[N]$ in the sense that the number of indices of $[N]$ that are neither in $\mathcal{H}_X$ nor in $\mathcal{L}_X$ is $o(N)$.

Let $\mathcal{L}_X^{\mathrm{c}}$ be the complement of $\mathcal{L}_X$, and note that

$$\lim_{N\to\infty} \frac{1}{N}|\mathcal{L}_X^{\mathrm{c}}| = H(X). \tag{4.12}$$

In addition, given $\{U_i\}_{i\in\mathcal{L}_X^{\mathrm{c}}}$, we can recover the whole vector $U_{1:N}$ in a successive manner, since U_i is approximately a deterministic function of $U_{1:i-1}$ for $i \in \mathcal{L}_X$. Consequently, we can compress $X_{1:N}$ into the sequence $\{U_i\}_{i\in\mathcal{L}_X^{\mathrm{c}}}$ that has a size roughly of $NH(X)$.

Encoding. Given the vector $x_{1:N}$ that we want to compress, the encoder computes $u_{1:N} = x_{1:N}G_N$ and outputs the values of $u_{1:N}$ in the positions $\mathcal{L}_X^{\mathrm{c}}$, i.e., it outputs $\{u_i\}_{i\in\mathcal{L}_X^{\mathrm{c}}}$.

Decoding. The decoder receives $\{u_i\}_{i\in\mathcal{L}_X^{\mathrm{c}}}$ and computes an estimate $\hat{u}_{1:N}$ of $u_{1:N}$ using the rule

$$\hat{u}_i = \begin{cases} u_i, & \text{if } i \in \mathcal{L}_X^{\mathrm{c}} \\ \arg\max\limits_{u\in\{0,1\}} \mathbb{P}_{U_i|U_{1:i-1}}(u \mid u_{1:i-1}), & \text{if } i \in \mathcal{L}_X \end{cases}. \tag{4.13}$$

Note that the conditional probabilities $\mathbb{P}_{U_i|U_{1:i-1}}(u \mid u_{1:i-1})$, for $u \in \{0,1\}$, can be computed recursively with complexity $\Theta(N \log_2 N)$.

Performance. As explained above, for $i \in \mathcal{L}_X$, the bit U_i is almost deterministic given its past $U_{1:i-1}$. Therefore, for $i \in \mathcal{L}_X$, the distribution $\mathbb{P}_{U_i|U_{1:i-1}}(u \mid u_{1:i-1})$ is highly biased towards the correct value u_i. Indeed, the block error probability P_{B} can be upper bounded by

$$P_{\mathrm{B}} = \mathbb{P}(\hat{U}_{1:N} \neq U_{1:N}) \leq \sum_{i\in\mathcal{L}_X} Z(U_i \mid U_{1:i-1}) = O(2^{-N^\beta}), \quad \forall\, \beta \in (0, 1/2). \tag{4.14}$$

Addition of Side Information. This is a slight extension of the previous case, and it is also discussed in [91]. Let $(X, Y) \sim p_{X,Y}$ be a pair of random variables, where we think of X as the source to be compressed and of Y as a *side information* about X. Given the vector $(X_{1:N}, Y_{1:N})$ of N independent samples from the

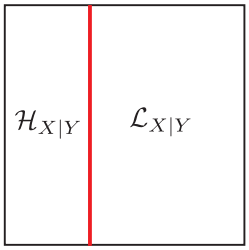


Figure 4.3 – A simple graphical representation of the sets $\mathcal{H}_{X|Y}$ and $\mathcal{L}_{X|Y}$ for the lossless compression scheme with side information. The whole square represents $[N]$. The sets $\mathcal{H}_{X|Y}$ and $\mathcal{L}_{X|Y}$ almost form a partition of $[N]$ in the sense that the number of indices of $[N]$ that are neither in $\mathcal{H}_{X|Y}$ nor in $\mathcal{L}_{X|Y}$ is $o(N)$.

distribution $p_{X,Y}$, the problem is to compress $X_{1:N}$ into a codeword of size roughly $NH(X \mid Y)$, so that the decoder is able to recover the whole vector $X_{1:N}$ by using the codeword and the side information $Y_{1:N}$.

Define $U_{1:N} = X_{1:N}G_N$ and consider the sets

$$\mathcal{H}_{X|Y} = \{i \in [N] \colon Z(U_i \mid U_{1:i-1}, Y_{1:N}) \geq 1 - \delta_N\}, \tag{4.15}$$

representing the positions such that U_i is approximately uniformly distributed and independent of $(U_{1:i-1}, Y_{1:N})$, and

$$\mathcal{L}_{X|Y} = \{i \in [N] \colon Z(U_i \mid U_{1:i-1}, Y_{1:N}) \leq \delta_N\}, \tag{4.16}$$

representing the positions such that U_i is approximately a deterministic function of $(U_{1:i-1}, Y_{1:N})$. The situation is schematically represented in Figure 4.3.

Note that lossless compression without side information can be considered as lossless compression with side information $\tilde{Y}$, where $\tilde{Y}$ is independent of X (say, e.g., that $\tilde{Y}$ is constant). Therefore, $\tilde{Y}$ does not add any information about X and it can be regarded as a degraded version of Y. Therefore, the following inclusion relations hold:

$$\begin{aligned} \mathcal{H}_{X|Y} &\subseteq \mathcal{H}_X, \\ \mathcal{L}_X &\subseteq \mathcal{L}_{X|Y}. \end{aligned} \tag{4.17}$$

A relationship analogous to (4.11) holds, namely,

$$\begin{aligned} \lim_{N\to\infty} \frac{1}{N} |\mathcal{H}_{X|Y}| &= H(X \mid Y), \\ \lim_{N\to\infty} \frac{1}{N} |\mathcal{L}_{X|Y}| &= 1 - H(X \mid Y). \end{aligned} \tag{4.18}$$

Given a realization of $X_{1:N}$, namely $x_{1:N}$, the encoder constructs $u_{1:N} = x_{1:N}G_N$ and outputs $\{u_i\}_{i\in\mathcal{L}^{\mathrm{c}}_{X|Y}}$ as the compressed version of $x_{1:N}$. The decoder, using the side information $y_{1:N}$ and a decoding rule similar to (4.13), is able to reconstruct $x_{1:N}$ reliably with vanishing block error probability.

Note that we have already defined[2] the sets $\mathcal{H}_{X|Y}$ and $\mathcal{L}_{X|Y}$ in Section 1.3.2, when describing how to achieve the capacity of a binary memoryless *symmetric*

[2]Compare (4.15), (4.16), and (4.18) with (1.15) and (1.17).

channel with polar codes. Hence, it is not surprising that we can use ideas from lossless compression to achieve the capacity of a general (hence, possibly asymmetric) binary memoryless channel.

4.4.2 Transmission over Binary Memoryless Channels

Problem Statement. Let W be a binary memoryless channel with input X and output Y. Fix a distribution p_X for the random variable X. The aim is to transmit over W with a rate close to $I(X;Y)$.

Design of the Scheme. Let $U_{1:N} = X_{1:N}G_N$, where $X_{1:N}$ is a vector of N i.i.d. components drawn according to p_X. Consider the sets $\mathcal{H}_X$ and $\mathcal{L}_X$ defined in (4.10). From the discussion about lossless compression, we know that, for $i \in \mathcal{H}_X$, the bit U_i is approximately uniformly distributed and independent of $U_{1:i-1}$ and that, for $i \in \mathcal{L}_X$, the bit U_i is approximately a deterministic function of the past $U_{1:i-1}$. Now, assume that the channel output $Y_{1:N}$ is given, and interpret this as side information on $X_{1:N}$. Consider the sets $\mathcal{H}_{X|Y}$ and $\mathcal{L}_{X|Y}$ as defined in (4.15) and (4.16), respectively. To recall, for $i \in \mathcal{H}_{X|Y}$, U_i is approximately uniformly distributed and independent of $(U_{1:i-1}, Y_{1:N})$ and, for $i \in \mathcal{L}_{X|Y}$, U_i becomes approximately a deterministic function of $(U_{1:i-1}, Y_{1:N})$.

To construct a polar code for the channel W, we proceed as follows. We place the information in the positions indexed by $\mathcal{I} = \mathcal{H}_X \cap \mathcal{L}_{X|Y}$. Indeed, if $i \in \mathcal{I}$, then U_i is approximately uniformly distributed given $U_{1:i-1}$, as $i \in \mathcal{H}_X$. This implies that U_i is suitable for containing information. Furthermore, U_i is approximately a deterministic function if we are given $U_{1:i-1}$ *and* $Y_{1:N}$, as $i \in \mathcal{L}_{X|Y}$. This implies that it is also decodable in a successive manner, given the channel output. Furthermore, we have that

$$\begin{aligned} \lim_{N\to\infty} \frac{1}{N} |\mathcal{I}| &\overset{\text{(a)}}{=} \lim_{N\to\infty} \frac{1}{N} |\mathcal{L}_{X|Y} \setminus \mathcal{L}_X| \\ &\overset{\text{(b)}}{=} \lim_{N\to\infty} \frac{1}{N} |\mathcal{L}_{X|Y}| - \lim_{N\to\infty} \frac{1}{N} |\mathcal{L}_X| \\ &\overset{\text{(c)}}{=} H(X) - H(X \mid Y) = I(X;Y), \end{aligned} \tag{4.19}$$

where the equality (a) uses that the number of indices in $[N]$ that are neither in $\mathcal{H}_X$ nor in $\mathcal{L}_X$ is $o(N)$; the equality (b) uses (4.17); and the equality (c) uses (4.11) and (4.18). As a result, our requirement on the transmission rate is met.

The remaining positions are frozen. More precisely, they are divided into two subsets, namely $\mathcal{F}_{\mathrm{r}} = \mathcal{H}_X \cap \mathcal{L}^{\mathrm{c}}_{X|Y}$ and $\mathcal{F}_{\mathrm{d}} = \mathcal{H}^{\mathrm{c}}_X$. For $i \in \mathcal{F}_{\mathrm{r}}$, U_i is independent of $U_{1:i-1}$, but cannot be reliably decoded using $Y_{1:N}$. Hence, we fill these positions with bits chosen uniformly at random, and this randomness is assumed to be shared between the transmitter and the receiver (i.e., the encoder and the decoder know the values associated with these positions). For $i \in \mathcal{F}_{\mathrm{d}}$, the value of U_i has to be chosen in a particular way. Indeed, almost all these positions are in $\mathcal{L}_X$, hence U_i is approximately a deterministic function of $U_{1:i-1}$. Below, we discuss in detail how to choose the values associated with the positions in $\mathcal{F}_{\mathrm{d}}$. The situation is schematically represented in Figure 4.4.

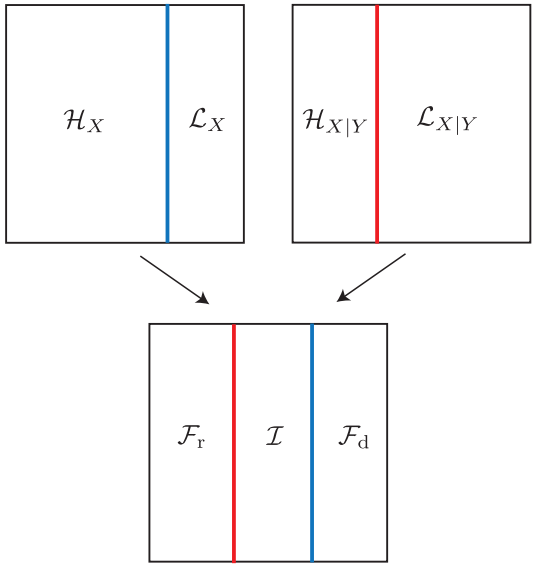


Figure 4.4 – Graphical representation of the sets associated with the channel coding problem. The two images on top represent how the set $[N]$ (the whole square) is partitioned by the source X (top left), and by the source X together with the output Y assumed as a side information (top right). Since $\mathcal{H}_{X|Y} \subseteq \mathcal{H}_X$ and $\mathcal{L}_X \subseteq \mathcal{L}_{X|Y}$, the set of indices $[N]$ can be partitioned into three subsets (bottom image): the information indices $\mathcal{I} = \mathcal{H}_X \cap \mathcal{L}_{X|Y}$; the frozen indices $\mathcal{F}_\mathrm{r} = \mathcal{H}_X \cap \mathcal{L}^\mathrm{c}_{X|Y}$ filled with bits chosen uniformly at random; and the frozen indices $\mathcal{F}_\mathrm{d} = \mathcal{H}^\mathrm{c}_X$ chosen according to either a "randomized rounding" rule or an "argmax" rule.

Encoding. We place the information into the positions indexed by $\mathcal{I}$, hence let $\{u_i\}_{i\in\mathcal{I}}$ denote the information bits to be transmitted. Then, we fill the positions indexed by $\mathcal{F}_\mathrm{r}$ with a random sequence that is shared between the transmitter and the receiver, hence let $\{u_i\}_{i\in\mathcal{F}_\mathrm{r}}$ be the particular realization of this sequence.

Let us now consider the encoding of the positions in $\mathcal{F}_\mathrm{d}$. An analogous problem was first considered in Section III of [93], where polar codes were used for lossy source coding. There are at least two possible approaches for solving this issue.

On the one hand, we can use a "randomized rounding" rule that consists of setting the value of bit i according to the distribution $\mathbb{P}_{U_i|U_{1:i-1}}$. In formulae,

$$u_i = \begin{cases} 0, & \text{w.p. } \mathbb{P}_{U_i|U_{1:i-1}}(0 \mid u_{1:i-1}) \\ 1, & \text{w.p. } \mathbb{P}_{U_i|U_{1:i-1}}(1 \mid u_{1:i-1}) \end{cases} \tag{4.20}$$

The random number generator used to construct this sequence is shared between the transmitter and the receiver. The "randomized rounding" rule yields a provable result [121].

On the other hand, we can use an "argmax" rule that consists of setting bit i to the value that maximizes $\mathbb{P}_{U_i|U_{1:i-1}}$. In formulae,

$$u_i = \arg \max_{u\in\{0,1\}} \mathbb{P}_{U_i|U_{1:i-1}}(u \mid u_{1:i-1}). \tag{4.21}$$

The "argmax" rule seems to perform slightly better in numerical simulations, but proving rigorous results under the "argmax" rule remains an open problem.

Eventually, the elements of $\{u_i\}_{i\in\mathcal{F}_\mathrm{d}}$ are computed in successive order by using either a "randomized rounding" or an "argmax" rule, and the probabilities $\mathbb{P}_{U_i|U_{1:i-1}}(u \mid u_{1:i-1})$ can be obtained recursively with complexity $\Theta(N\log_2 N)$. As $G_N = G_N^{(-1)}$, the vector $x_{1:N} = u_{1:N}G_N$ is transmitted over the channel.

Decoding. The decoder receives $y_{1:N}$ and computes the estimate $\hat{u}_{1:N}$ of $u_{1:N}$ according to the rule

$$\hat{u}_i = \begin{cases} u_i, & \text{if } i \in \mathcal{F}_\mathrm{r} \\ \arg\max\limits_{u\in\{0,1\}} \mathbb{P}_{U_i|U_{1:i-1}}(u \mid u_{1:i-1}), & \text{if } i \in \mathcal{F}_\mathrm{d} \\ \arg\max\limits_{u\in\{0,1\}} \mathbb{P}_{U_i|U_{1:i-1},Y_{1:N}}(u \mid u_{1:i-1}, y_{1:N}), & \text{if } i \in \mathcal{I} \end{cases} \tag{4.22}$$

where $\mathbb{P}_{U_i|U_{1:i-1},Y_{1:N}}(u \mid u_{1:i-1}, y_{1:N})$ can be computed recursively with complexity $\Theta(N\log_2 N)$. In (4.22), we assume that the "argmax" rule is used to encode the positions in $\mathcal{F}_\mathrm{d}$. If the "randomized rounding" rule is used, then the decoder can still correctly recover u_i, since the random sequence used in (4.20) is shared between the transmitter and the receiver.

Performance. The block error probability P_B can be upper bounded by

$$P_\mathrm{B} \overset{\text{(a)}}{\leq} \sum_{i\in\mathcal{I}} Z(U_i \mid U_{1:i-1}, Y_{1:N}) \overset{\text{(b)}}{=} O(2^{-N^\beta}), \quad \forall\, \beta \in (0, 1/2). \tag{4.23}$$

Let us briefly comment on how to obtain formula (4.23). The inequality (a) comes from the union bound: the error probability under SC decoding is upper bounded by the sum of the probabilities of making a mistake while decoding each of the information bits (the remaining bits are frozen, hence known at the decoder). Furthermore, the probability of decoding incorrectly the i-th synthetic channel ($i \in \mathcal{I}$) is upper bounded by its Bhattacharyya parameter, namely, $Z(U_i \mid U_{1:i-1}, Y_{1:N})$. Finally, the equality (b) comes from the definition of the set $\mathcal{I}$, that contains positions i such that $Z(U_i \mid U_{1:i-1}, Y_{1:N})$ is small enough.

4.5 Polar Codes for Superposition Region

The following theorem provides our main result about the achievability of Bergmans' superposition region for DM-BCs with polar codes (compare with Theorem 4.1).

Theorem 4.6 (Polar codes for Superposition Region)**.** *Consider a two-user DM-BC $p_{Y_1,Y_2|X}$ with a binary-input alphabet, where X denotes the input to the channel, and Y_1, Y_2 denote the outputs at the first and second receiver, respectively. Let V be an auxiliary binary random variable. Then, for any joint distribution $p_{V,X}$ such that $V - X - (Y_1, Y_2)$ forms a Markov chain and for any rate pair (R_1, R_2) satisfying the constraints in* (4.1)*, there exists a sequence of polar codes with an increasing block length N that achieves this rate pair with encoding and decoding complexity $\Theta(N\log_2 N)$ and a block error probability decaying like $O(2^{-N^\beta})$ for any $\beta \in (0, 1/2)$.*

Problem Statement. Let $(V, X) \sim p_{V,X} = p_V p_{X|V}$. We will show how to transmit over the two-user DM-BC $p_{Y_1,Y_2|X}$ achieving the rate pair

$$(R_1, R_2) = (I(X;Y_1) - I(V;Y_2), I(V;Y_2)), \tag{4.24}$$

when $I(V;Y_1) < I(V;Y_2) < I(X;Y_1)$. Once we have accomplished this, we will see that a slight modification of this scheme enables us to achieve, in addition, the rate pair

$$(R_1, R_2) = (I(X;Y_1 \mid V), \min_{l\in\{1,2\}} I(V;Y_l)). \tag{4.25}$$

Therefore, by Proposition 4.1, we can achieve the whole region (4.1) and Theorem 4.6 is proved. Note that if polar coding achieves the rate pairs (4.24) and (4.25) with complexity $\Theta(N \log_2 N)$ and a block error probability $O(2^{-N^\beta})$, then for any other rate pair in the region (4.1), there exists a sequence of polar codes with an increasing block length N whose complexity and block error probability have the same asymptotic scalings.

Design of the Scheme. Set $U_{1:N}^{(2)} = V_{1:N}G_N$. As in the case of the transmission over a general binary memoryless channel with V in place of X and Y_l ($l \in \{1,2\}$) in place of Y, define the sets $\mathcal{H}_V$, $\mathcal{L}_V$, $\mathcal{H}_{V|Y_l}$, and $\mathcal{L}_{V|Y_l}$, analogously to Section 4.4.2, as follows:

$$\begin{aligned}
\mathcal{H}_V &= \{i \in [N]\colon Z(U_i^{(2)} \mid U_{1:i-1}^{(2)}) \geq 1 - \delta_N\},\\
\mathcal{L}_V &= \{i \in [N]\colon Z(U_i^{(2)} \mid U_{1:i-1}^{(2)}) \leq \delta_N\},\\
\mathcal{H}_{V|Y_l} &= \{i \in [N]\colon Z(U_i^{(2)} \mid U_{1:i-1}^{(2)}, Y_{1:N}^{(l)}) \geq 1 - \delta_N\},\\
\mathcal{L}_{V|Y_l} &= \{i \in [N]\colon Z(U_i^{(2)} \mid U_{1:i-1}^{(2)}, Y_{1:N}^{(l)}) \leq \delta_N\},
\end{aligned} \tag{4.26}$$

where $Y_{1:N}^{(l)}$ denotes the vector of length N received by the l-th user. These sets satisfy, for $l \in \{1,2\}$,

$$\begin{aligned}
\lim_{N\to\infty} \frac{1}{N} |\mathcal{H}_V| &= H(V),\\
\lim_{N\to\infty} \frac{1}{N} |\mathcal{L}_V| &= 1 - H(V),\\
\lim_{N\to\infty} \frac{1}{N} |\mathcal{H}_{V|Y_l}| &= H(V \mid Y_l),\\
\lim_{N\to\infty} \frac{1}{N} |\mathcal{L}_{V|Y_l}| &= 1 - H(V \mid Y_l).
\end{aligned} \tag{4.27}$$

Set $U_{1:N}^{(1)} = X_{1:N}G_N$. By thinking of V as side information on X and by considering the transmission of X over the memoryless channel with output Y_1, define also the sets $\mathcal{H}_{X|V}$, $\mathcal{L}_{X|V}$, $\mathcal{H}_{X|V,Y_1}$, and $\mathcal{L}_{X|V,Y_1}$, as follows:

$$\begin{aligned}
\mathcal{H}_{X|V} &= \{i \in [N]\colon Z(U_i^{(1)} \mid U_{1:i-1}^{(1)}, V_{1:N}) \geq 1 - \delta_N\},\\
\mathcal{L}_{X|V} &= \{i \in [N]\colon Z(U_i^{(1)} \mid U_{1:i-1}^{(1)}, V_{1:N}) \leq \delta_N\},\\
\mathcal{H}_{X|V,Y_1} &= \{i \in [N]\colon Z(U_i^{(1)} \mid U_{1:i-1}^{(1)}, V_{1:N}, Y_{1:N}^{(1)}) \geq 1 - \delta_N\},\\
\mathcal{L}_{X|V,Y_1} &= \{i \in [N]\colon Z(U_i^{(1)} \mid U_{1:i-1}^{(1)}, V_{1:N}, Y_{1:N}^{(1)}) \leq \delta_N\},
\end{aligned} \tag{4.28}$$

which satisfy

$$\begin{aligned}\lim_{N\to\infty}\frac{1}{N}\left|\mathcal{H}_{X|V}\right| &= H(X \mid V),\\ \lim_{N\to\infty}\frac{1}{N}\left|\mathcal{L}_{X|V}\right| &= 1-H(X \mid V),\\ \lim_{N\to\infty}\frac{1}{N}\left|\mathcal{H}_{X|V,Y_1}\right| &= H(X \mid V,Y_1),\\ \lim_{N\to\infty}\frac{1}{N}\left|\mathcal{L}_{X|V,Y_1}\right| &= 1-H(X \mid V,Y_1).\end{aligned} \tag{4.29}$$

First, consider only the point-to-point communication problem between the transmitter and the second receiver. As discussed in Section 4.4.2, for this scenario, the correct choice is to place the information bits in those positions of $U_{1:N}^{(2)}$ that are indexed by the set $\mathcal{I}^{(2)} = \mathcal{H}_V \cap \mathcal{L}_{V|Y_2}$. If, in addition, we restrict ourselves to the positions in $\mathcal{I}^{(2)}$ that are contained in $\mathcal{I}_v^{(1)} = \mathcal{H}_V \cap \mathcal{L}_{V|Y_1}$, also the first receiver will be able to decode this message. Indeed, recall that in the superposition coding scheme, before decoding its own message, the first receiver needs to decode the message intended for the second receiver. Consequently, for sufficiently large N, the first receiver knows the vector $U_{1:N}^{(2)}$ with high probability, hence also the vector $V_{1:N} = U_{1:N}^{(2)} G_N$ (recall that $G_N^{-1} = G_N$).

Now, consider the point-to-point communication problem between the transmitter and the first receiver, given the side information $V_{1:N}$ (following our discussion, as we let N tend to infinity, the vector $V_{1:N}$ is known to the first receiver with probability that tends to 1). From Section 4.4.2, we know that the information has to be placed in those positions of $U_{1:N}^{(1)}$ that are indexed by $\mathcal{I}^{(1)} = \mathcal{H}_{X|V} \cap \mathcal{L}_{X|V,Y_1}$.

The cardinalities of these information sets are given by

$$\begin{aligned}\lim_{N\to\infty}\frac{1}{N}\left|\mathcal{I}^{(2)}\right| &= I(V;Y_2),\\ \lim_{N\to\infty}\frac{1}{N}\left|\mathcal{I}_v^{(1)}\right| &= I(V;Y_1),\\ \lim_{N\to\infty}\frac{1}{N}\left|\mathcal{I}^{(1)}\right| &= I(X;Y_1 \mid V).\end{aligned} \tag{4.30}$$

Let us now get back to the broadcasting scenario and see how the previous observations can be used to construct a polar coding scheme. Recall that $X_{1:N}$ is transmitted over the channel, the second receiver only decodes its intended message, but the first receiver decodes both messages.

We begin by reviewing the AGG scheme [100]. This scheme achieves the rate pair

$$(R_1, R_2) = (I(X;Y_1 \mid V), I(V;Y_2)), \tag{4.31}$$

assuming that $p_{Y_1|V} \succ p_{Y_2|V}$. Under this assumption, we have $\mathcal{L}_{V|Y_2} \subseteq \mathcal{L}_{V|Y_1}$ and therefore $\mathcal{I}^{(2)} \subseteq \mathcal{I}_v^{(1)}$. Consequently, we can in fact use the point-to-point solutions outlined above, i.e., the second user can place his information in $\mathcal{I}^{(2)}$ and decode, and the first user will also be able to decode this message. Furthermore, once the message intended for the second user is known by the first user, the latter can decode his own information placed in the positions of $\mathcal{I}^{(1)}$.

Let us now see how to eliminate the restriction imposed by the degradation condition $p_{Y_1|V} \succ p_{Y_2|V}$. Recall that we want to achieve the rate pair (4.24) when $I(V;Y_1) < I(V;Y_2) < I(X;Y_1)$. The set of indices of the information bits for the first user is exactly the same as before, namely the positions of $U_{1:N}^{(1)}$ indexed by $\mathcal{I}^{(1)}$. The only difficulty lies in designing a coding scheme in which *both* receivers can decode the message intended for the second user.

First of all, observe that we can use all the positions in $\mathcal{I}_v^{(1)} \cap \mathcal{I}^{(2)}$, since they are decodable by both users. Let us define

$$\mathcal{D}^{(2)} = \mathcal{I}^{(2)} \setminus \mathcal{I}_v^{(1)}. \tag{4.32}$$

If $p_{Y_1|V} \succ p_{Y_2|V}$, as before, then $\mathcal{D}^{(2)} = \emptyset$ (i.e., all the positions decodable by the second user are decodable also by the first user). However, in the general case, where it is no longer assumed that $p_{Y_1|V} \succ p_{Y_2|V}$, the set $\mathcal{D}^{(2)}$ is not empty and those positions cannot be decoded by the first user.

Note that there is a similar set, but with the roles of the two users exchanged, call it $\mathcal{D}^{(1)}$, namely,

$$\mathcal{D}^{(1)} = \mathcal{I}_v^{(1)} \setminus \mathcal{I}^{(2)}. \tag{4.33}$$

The set $\mathcal{D}^{(1)}$ contains the positions of $U_{1:N}^{(2)}$ that are decodable by the first user, but not by the second user. Observe further that $|\mathcal{D}^{(1)}| \leq |\mathcal{D}^{(2)}|$ for sufficiently large N. Indeed, since the equality

$$|A \setminus B| - |B \setminus A| = |A| - |B| \tag{4.34}$$

holds for any two finite sets A and B, it follows from (4.30)–(4.32) that for sufficiently large N

$$\frac{1}{N}\left(|\mathcal{D}^{(2)}| - |\mathcal{D}^{(1)}|\right) = \frac{1}{N}\left(|\mathcal{I}^{(2)}| - |\mathcal{I}_v^{(1)}|\right) \tag{4.35}$$

$$= I(V;Y_2) - I(V;Y_1) + o(1) \geq 0. \tag{4.36}$$

Assume at first that the two sets are of equal size. The general case will require only a small modification.

Now, the idea is to consider the "chaining" construction introduced in [48] in the context of universal polar codes. Recall that we are only interested in the message intended for the second user, but that both receivers must be able to decode this message. Our scheme consists in transmitting k polar blocks, and in repeating ("chaining") some information. More precisely, in block 1 fill the positions indexed by $\mathcal{D}^{(1)}$ with information, but set the bits indexed by $\mathcal{D}^{(2)}$ to a fixed known sequence. In block j ($j \in \{2, \cdots, k-1\}$), fill the positions indexed by $\mathcal{D}^{(1)}$ again with information, and repeat the bits contained in the positions indexed by $\mathcal{D}^{(1)}$ of block $j-1$ into the positions indexed by $\mathcal{D}^{(2)}$ of block j. In the final block k, put a known sequence in the positions indexed by $\mathcal{D}^{(1)}$, and repeat in the positions indexed by $\mathcal{D}^{(2)}$ the bits in the positions indexed by $\mathcal{D}^{(1)}$ of block $k-1$. The remaining bits are frozen and, as in Section 4.4.2, they are divided into the two subsets $\mathcal{F}_{\mathrm{d}}^{(2)} = \mathcal{H}_V^{\mathrm{c}}$ and $\mathcal{F}_{\mathrm{r}}^{(2)} = \mathcal{H}_V \cap \mathcal{L}_{V|Y_2}^{\mathrm{c}} \subset \mathcal{H}_V$. In the first case, $U_i^{(2)}$ is approximately a deterministic function of $U_{1:i-1}^{(2)}$, whereas in the second case $U_i^{(2)}$ is approximately independent of $U_{1:i-1}^{(2)}$.

Note that we lose some rate, as at the boundary we put a known sequence into some bits that were supposed to contain information. However, this rate loss decays like $1/k$. Hence, by choosing a sufficiently large k, we can achieve a rate that is arbitrarily close to the intended rate.

We claim that in the above construction both users can decode all blocks, but the first receiver has to decode "forward", starting with block 1 and ending with block k, whereas the second receiver decodes "backwards", starting with block k and ending with block 1. Let us discuss this procedure in some more detail. Look at the first user and start with block 1. By construction, information is only contained in the positions indexed by $\mathcal{D}^{(1)}$ as well as $\mathcal{I}_v^{(1)} \cap \mathcal{I}^{(2)}$, whereas the positions indexed by $\mathcal{D}^{(2)}$ are set to known values. Hence, the first user can decode this block. For block j ($j \in \{2, \cdots, k-1\}$), the situation is similar: the first user decodes the positions indexed by $\mathcal{D}^{(1)}$ and $\mathcal{I}_v^{(1)} \cap \mathcal{I}^{(2)}$, whereas the positions in $\mathcal{D}^{(2)}$ contain repeated information that has been already decoded in the previous block. An analogous analysis applies to block k, in which the positions indexed by $\mathcal{D}^{(1)}$ are also fixed to a known sequence. The second user proceeds exactly in the same fashion, but goes backwards.

To get to the general case, we need to discuss what happens when $|\mathcal{D}^{(1)}| < |\mathcal{D}^{(2)}|$ (due to (4.35), in general $|\mathcal{D}^{(1)}| \leq |\mathcal{D}^{(2)}|$ for sufficiently large N, but the special case where the two sets are of equal size has been already addressed). In this case, we do not have sufficiently many positions in $\mathcal{D}^{(1)}$ to repeat all the information contained in $\mathcal{D}^{(2)}$. To get around this problem, we pick sufficiently many extra positions out of the vector $U_{1:N}^{(1)}$ indexed by $\mathcal{I}^{(1)}$, and repeat the extra information there.

In order to specify this scheme, let us introduce some notation for the various sets. Recall that we "chain" the positions in $\mathcal{D}^{(1)}$ with an equal amount of positions in $\mathcal{D}^{(2)}$. It does not matter what subset of $\mathcal{D}^{(2)}$ we pick, but call the chosen subset $\mathcal{R}^{(2)}$. Now, we still have some positions left in $\mathcal{D}^{(2)}$, call them $\mathcal{B}^{(2)}$. More precisely, $\mathcal{B}^{(2)} = \mathcal{D}^{(2)} \setminus \mathcal{R}^{(2)}$. Since $\mathcal{R}^{(2)} \subseteq \mathcal{D}^{(2)}$ and $|\mathcal{R}^{(2)}| = |\mathcal{D}^{(1)}|$, it follows from (4.35) that

$$\begin{aligned} \frac{1}{N}\,|\mathcal{B}^{(2)}| &= \frac{1}{N}\,(|\mathcal{D}^{(2)}| - |\mathcal{R}^{(2)}|) \\ &= \frac{1}{N}\,(|\mathcal{D}^{(2)}| - |\mathcal{D}^{(1)}|) \\ &= I(V;Y_2) - I(V;Y_1) + o(1) \geq 0. \end{aligned} \tag{4.37}$$

Let $\mathcal{B}^{(1)}$ be a subset of $\mathcal{I}^{(1)}$ such that $|\mathcal{B}^{(1)}| = |\mathcal{B}^{(2)}|$. Again, it does not matter what subset we pick. The existence of such a set $\mathcal{B}^{(1)}$, for sufficiently large N, is ensured by noticing that from (4.30), (4.37) and the Markovity of the chain $V - X - Y_1$ we obtain

$$\begin{aligned} \frac{1}{N}(|\mathcal{I}^{(1)}| - |\mathcal{B}^{(2)}|) &= I(X;Y_1 \mid V) - I(V;Y_2) + I(V;Y_1) + o(1) \\ &= I(X;Y_1) - I(V;Y_2) + o(1) \geq 0. \end{aligned} \tag{4.38}$$

Indeed, recall that we need to achieve the rate pair (4.24) when $I(V;Y_1) < I(V;Y_2) < I(X;Y_1)$.

As explained above, we place in $\mathcal{B}^{(1)}$ the value of those extra bits from $\mathcal{D}^{(2)}$ that will help the first user to decode the message of the second user in the next block.

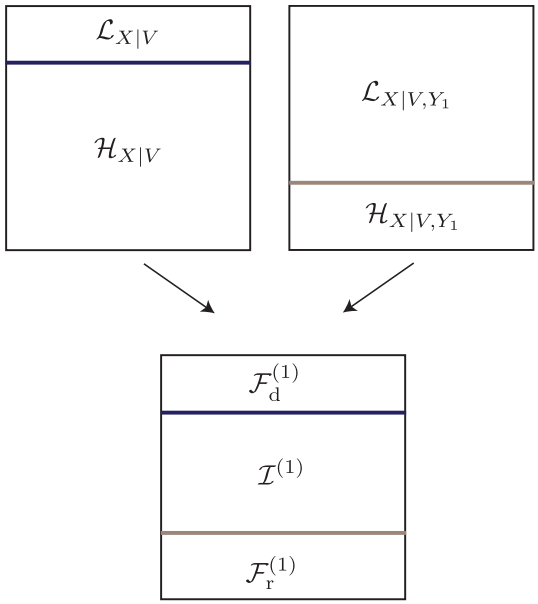


Figure 4.5 – Graphical representation of the sets associated with the first user for the superposition scheme. The set $[N]$ is partitioned into three subsets: the information indices $\mathcal{I}^{(1)}$; the frozen indices $\mathcal{F}_{\mathrm{r}}^{(1)}$ filled with bits chosen uniformly at random; the frozen indices $\mathcal{F}_{\mathrm{d}}^{(1)}$ chosen according to either a "randomized rounding" rule or an "argmax" rule.

Operationally, we repeat the information contained in the positions indexed by $\mathcal{B}^{(2)}$ into the positions indexed by $\mathcal{B}^{(1)}$ of the previous block. By doing this, the first user pays a rate penalty of $I(V;Y_2) - I(V;Y_1) + o(1)$ compared to his original rate given by $\frac{1}{N}|\mathcal{I}^{(1)}| = I(X;Y_1|V) + o(1)$.

To summarize, the first user puts information bits at positions $\mathcal{I}^{(1)} \setminus \mathcal{B}^{(1)}$, repeats in $\mathcal{B}^{(1)}$ the information bits in $\mathcal{B}^{(2)}$ for the next block, and freezes the rest. In the last block, the information set is the whole $\mathcal{I}^{(1)}$. The frozen positions are divided into the usual two subsets $\mathcal{F}_{\mathrm{r}}^{(1)} = \mathcal{H}_{X|V} \cap \mathcal{L}_{X|V,Y_1}^{\mathrm{c}}$ and $\mathcal{F}_{\mathrm{d}}^{(1)} = \mathcal{H}_{X|V}^{\mathrm{c}}$ that contain positions such that $U_i^{(1)}$ is or is not, respectively, approximately independent of $(U_{1:i-1}^{(1)}, V_{1:N})$. The situation is schematically represented in Figures 4.5–4.7.

Suppose that, by applying the same scheme with $k \to \infty$, we let $\frac{1}{N}|\mathcal{B}^{(2)}|$ shrink from $I(V;Y_2) - I(V;Y_1) + o(1)$ in (4.37) to $o(1)$. Then, we obtain the whole line going from the rate pair $(I(X;Y_1) - I(V;Y_2), I(V;Y_2))$ to $(I(X;Y_1 \mid V), I(V;Y_1))$ without time-sharing.[3]

Finally, in order to obtain the rate pair $(I(X;Y_1 \mid V), I(V;Y_2))$ when $I(V;Y_2) \leq I(V;Y_1)$, it suffices to consider the case where $\mathcal{B}^{(2)} = \emptyset$ and switch the roles of $\mathcal{I}^{(2)}$ and $\mathcal{I}_v^{(1)}$ in the discussion concerning the second user.

[3]The reader will be able to verify this property by relying on (4.41) and (4.44); this property is mentioned, however, at this stage as part of the exposition of the polar coding scheme.

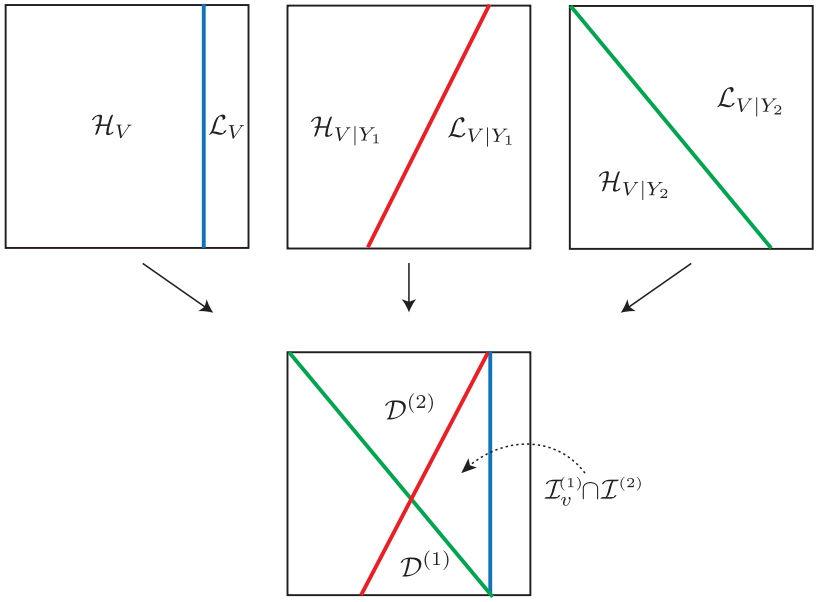


Figure 4.6 – Graphical representation of the sets associated with the second user for the superposition scheme: $\mathcal{I}_v^{(1)} \cap \mathcal{I}^{(2)}$ contains the indices that are decodable by both users; $\mathcal{D}^{(1)} = \mathcal{I}_v^{(1)} \setminus \mathcal{I}^{(2)}$ contains the indices that are decodable by the first user, but not by the second user; $\mathcal{D}^{(2)} = \mathcal{I}^{(2)} \setminus \mathcal{I}_v^{(1)}$ contains the indices that are decodable by the second user, but not by the first user.

Encoding. Let us start from the second user, and encode block by block. For block 1:

- Let $\{u_i^{(2)}\}_{i \in \mathcal{I}_v^{(1)}}$ denote the information bits.
- Denote by $\{u_i^{(2)}\}_{i \in \mathcal{F}_{\mathrm{r}}^{(2)}}$ a particular realization of a random sequence that is shared between the transmitter and both receivers.
- As discussed in Section 4.4.2, for $i \in \mathcal{F}_{\mathrm{d}}^{(2)}$, we can either use a "randomized rounding" rule, i.e.,

$$u_i^{(2)} = \begin{cases} 0, & \text{w.p. } \mathbb{P}_{U_i^{(2)}|U_{1:i-1}^{(2)}}(0 \mid u_{1:i-1}^{(2)}) \\ 1, & \text{w.p. } \mathbb{P}_{U_i^{(2)}|U_{1:i-1}^{(2)}}(1 \mid u_{1:i-1}^{(2)}) \end{cases} \tag{4.39}$$

  or an "argmax" rule, i.e.,

$$u_i^{(2)} = \arg \max_{u \in \{0,1\}} \mathbb{P}_{U_i^{(2)}|U_{1:i-1}^{(2)}}(u \mid u_{1:i-1}^{(2)}). \tag{4.40}$$

  In case the "randomized rounding" rule is employed, the random sequence in (4.39) is shared between the transmitter and the receiver.

For block j ($j \in \{2, \cdots, k-1\}$):

- Let $\{u_i^{(2)}\}_{i \in \mathcal{I}_v^{(1)} \cup \mathcal{B}^{(2)}}$ denote the information bits.

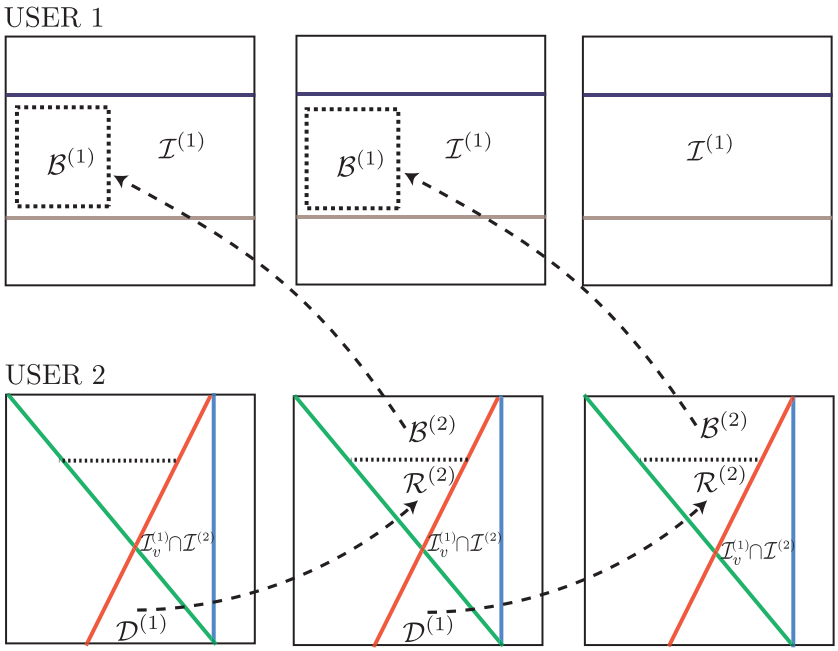


Figure 4.7 – Graphical representation of the chaining construction for the superposition scheme with $k = 3$: the set $\mathcal{D}^{(1)}$ is repeated into the set $\mathcal{R}^{(2)}$ of the following block; the set $\mathcal{B}^{(2)}$ is repeated into the set $\mathcal{B}^{(1)}$ of the previous block (belonging to a different user).

- $\{u_i^{(2)}\}_{i \in \mathcal{R}^{(2)}}$ contains a copy of the sequence $\{u_i^{(2)}\}_{i \in \mathcal{D}^{(1)}}$ of block $j-1$.
- The frozen sequences $\{u_i^{(2)}\}_{i \in \mathcal{F}_{\mathrm{r}}^{(2)}}$ and $\{u_i^{(2)}\}_{i \in \mathcal{F}_{\mathrm{d}}^{(2)}}$ are chosen as in block 1.

For block k (the last one):

- Let $\{u_i^{(2)}\}_{i \in (\mathcal{I}_v^{(1)} \cap \mathcal{I}^{(2)}) \cup \mathcal{B}^{(2)}}$ denote the information bits.
- $\{u_i^{(2)}\}_{i \in \mathcal{R}^{(2)}}$ contains the sequence $\{u_i^{(2)}\}_{i \in \mathcal{D}^{(1)}}$ of block $k-1$.
- The frozen bits are chosen with the usual rules.

The rate of the second user is given by

$$\begin{aligned} R_2 &= \frac{1}{kN}\left[\left|\mathcal{I}_v^{(1)}\right| + (k-2)\left|\mathcal{I}_v^{(1)} \cup \mathcal{B}^{(2)}\right| + \left|(\mathcal{I}_v^{(1)} \cap \mathcal{I}^{(2)}) \cup \mathcal{B}^{(2)}\right|\right] \\ &= \left(\frac{k-1}{k}\right) I(V;Y_2) + \frac{1}{kN}\left|\mathcal{I}_v^{(1)} \cap \mathcal{I}^{(2)}\right| + o(1), \end{aligned} \tag{4.41}$$

where the second equality follows from (4.30) and (4.37), and from the fact that the sets $\mathcal{I}_v^{(1)}$ and $\mathcal{B}^{(2)}$ are disjoint. From (4.41), we obtain that, as k tends to infinity, R_2 approaches the required rate $I(V;Y_2)$. Then, the vector $v_{1:N} = u_{1:N}^{(2)} G_N$ is obtained.

The encoder for the first user knows $v_{1:N}$ and proceeds block by block:

- Let $\{u_i^{(1)}\}_{i\in\mathcal{I}^{(1)}\setminus\mathcal{B}^{(1)}}$ denote the information bits, except for block k, in which the information sequence is $\{u_i^{(1)}\}_{i\in\mathcal{I}^{(1)}}$.
- For block j ($j \in \{1,\cdots,k-1\}$), $\{u_i^{(1)}\}_{i\in\mathcal{B}^{(1)}}$ contains a copy of the sequence $\{u_i^{(2)}\}_{i\in\mathcal{B}^{(2)}}$ in block $j+1$.
- The frozen sequence $\{u_i^{(1)}\}_{i\in\mathcal{F}_{\mathrm{r}}^{(1)}}$ contains a realization of a random sequence shared between the encoder and the first decoder.
- As discussed in Section 4.4.2, for $i \in \mathcal{F}_{\mathrm{d}}^{(1)}$, we can either use a "randomized rounding" rule, i.e.,

$$u_i^{(1)} = \begin{cases} 0, & \text{w.p. } \mathbb{P}_{U_i^{(1)}|U_{1:i-1}^{(1)},V_{1:N}}(0 \mid u_{1:i-1}^{(1)}, v_{1:N}) \\ 1, & \text{w.p. } \mathbb{P}_{U_i^{(1)}|U_{1:i-1}^{(1)},V_{1:N}}(1 \mid u_{1:i-1}^{(1)}, v_{1:N}) \end{cases} \tag{4.42}$$

  or an "argmax" rule, i.e.,

$$u_i^{(1)} = \arg\max_{u\in\{0,1\}} \mathbb{P}_{U_i^{(1)}|U_{1:i-1}^{(1)},V_{1:N}}(u \mid u_{1:i-1}^{(1)}, v_{1:N}). \tag{4.43}$$

  In case the "randomized rounding" rule is employed, the random sequence in (4.42) is shared between the transmitter and the receiver.

The rate of the first user is given by

$$\begin{aligned} R_1 &= \frac{1}{kN}\left[(k-1)|\mathcal{I}^{(1)}\setminus\mathcal{B}^{(1)}| + |\mathcal{I}^{(1)}|\right] \\ &= I(X;Y_1 \mid V) - \frac{k-1}{k}\left(I(V;Y_2) - I(V;Y_1)\right) + o(1), \end{aligned} \tag{4.44}$$

where we use (4.30), (4.38), and the fact that $\mathcal{B}^{(1)}$ is a subset of $\mathcal{I}^{(1)}$ such that $|\mathcal{B}^{(1)}| = |\mathcal{B}^{(2)}|$. From (4.44), we obtain that, as k tends to infinity, R_1 approaches the required rate $I(X;Y_1) - I(V;Y_2)$. Finally, the vector $x_{1:N} = u_{1:N}^{(1)} G_N$ is transmitted over the channel. The encoding complexity per block is $\Theta(N \log_2 N)$.

Decoding. Let us start from the first user that receives the channel output $y_{1:N}^{(1)}$. The decoder acts block by block and reconstructs first $u_{1:N}^{(2)}$, computes $v_{1:N} = u_{1:N}^{(2)} G_N$, and then decodes $u_{1:N}^{(1)}$, thus recovering his own message. For block 1, the decision rule is given by

$$\hat{u}_i^{(2)} = \begin{cases} u_i^{(2)}, & \text{if } i \in \mathcal{F}_{\mathrm{r}}^{(2)} \\ \arg\max_{u\in\{0,1\}} \mathbb{P}_{U_i^{(2)}|U_{1:i-1}^{(2)}}(u \mid u_{1:i-1}^{(2)}), & \text{if } i \in \mathcal{F}_{\mathrm{d}}^{(2)} \\ \arg\max_{u\in\{0,1\}} \mathbb{P}_{U_i^{(2)}|U_{1:i-1}^{(2)},Y_{1:N}^{(1)}}(u \mid u_{1:i-1}^{(2)}, y_{1:N}^{(1)}), & \text{if } i \in \mathcal{I}_v^{(1)} \end{cases} \tag{4.45}$$

and

$$\hat{u}_i^{(1)} = \begin{cases} u_i^{(1)}, & \text{if } i \in \mathcal{F}_{\mathrm{r}}^{(1)} \\ \arg\max_{u\in\{0,1\}} \mathbb{P}_{U_i^{(1)}|U_{1:i-1}^{(1)},V_{1:N}}(u \mid u_{1:i-1}^{(1)}, v_{1:N}), & \text{if } i \in \mathcal{F}_{\mathrm{d}}^{(1)} \\ \arg\max_{u\in\{0,1\}} \mathbb{P}_{U_i^{(1)}|U_{1:i-1}^{(1)},V_{1:N},Y_{1:N}^{(1)}}(u \mid u_{1:i-1}^{(1)}, v_{1:N}, y_{1:N}^{(1)}), & \text{if } i \in \mathcal{I}^{(1)} \end{cases} \tag{4.46}$$

where we assume that the "argmax" rule is employed to encode the positions in $\mathcal{F}_{\mathrm{d}}^{(1)}$ and $\mathcal{F}_{\mathrm{d}}^{(2)}$. If the "randomized rounding" rule is used, then the decoder can still correctly recover $u_i^{(1)}$ and $u_i^{(2)}$, since the random sequences used in (4.39) and (4.42) are shared between the transmitter and the receiver.
For block j ($j \in \{2, \cdots, k-1\}$):

- $\{\hat{u}_i^{(2)}\}_{i \in \mathcal{B}^{(2)}}$ is deduced from $\{\hat{u}_i^{(1)}\}_{i \in \mathcal{B}^{(1)}}$ of block $j-1$.
- $\{\hat{u}_i^{(2)}\}_{i \in \mathcal{R}^{(2)}}$ is deduced from $\{\hat{u}_i^{(2)}\}_{i \in \mathcal{D}^{(1)}}$ of block $j-1$.
- For the remaining positions of $\hat{u}_i^{(2)}$, the decoding follows the rule in (4.45).
- The decoding of $\hat{u}_i^{(1)}$ proceeds as in (4.46).

This decoding rule works also for block k, with the only difference that the frozen set $\mathcal{F}_{\mathrm{r}}^{(2)}$ is bigger, and $\hat{u}_i^{(2)} = \arg\max_{u \in \{0,1\}} \mathbb{P}_{U_i^{(2)}|U_{1:i-1}^{(2)}, Y_{1:N}^{(1)}}(u \mid u_{1:i-1}^{(2)}, y_{1:N}^{(1)})$ only for $i \in \mathcal{I}_v^{(1)} \cap \mathcal{I}^{(2)}$.

Let us consider now the second user that reconstructs $u_{1:N}^{(2)}$ from the channel output $y_{1:N}^{(2)}$. As explained before, the decoding goes "backwards", starting from block k and ending with block 1. For block k, the decision rule is given by

$$\hat{u}_i^{(2)} = \begin{cases} u_i^{(2)}, & \text{if } i \in \mathcal{F}_{\mathrm{r}}^{(2)} \\ \arg\max\limits_{u \in \{0,1\}} \mathbb{P}_{U_i^{(2)}|U_{1:i-1}^{(2)}}(u \mid u_{1:i-1}^{(2)}), & \text{if } i \in \mathcal{F}_{\mathrm{d}}^{(2)} \\ \arg\max\limits_{u \in \{0,1\}} \mathbb{P}_{U_i^{(2)}|U_{1:i-1}^{(2)}, Y_{1:N}^{(2)}}(u \mid u_{1:i-1}^{(2)}, y_{1:N}^{(2)}), & \text{if } i \in (\mathcal{I}^{(1)} \cap \mathcal{I}^{(2)}) \cup \mathcal{R}^{(2)} \cup \mathcal{B}^{(2)} \end{cases} \tag{4.47}$$

Once again, we assume that the "argmax" rule is employed to encode the positions in $\mathcal{F}_{\mathrm{d}}^{(2)}$. Clearly, correct recovery is also possible under the "randomized rounding" rule.
For block j ($j \in \{2, \cdots, k-1\}$), the decoder recovers $\{u_i^{(2)}\}_{i \in \mathcal{D}^{(1)}}$ from $\{u_i^{(2)}\}_{i \in \mathcal{R}^{(2)}}$ of block $j+1$; for the remaining positions, the decision rule in (4.47) is used.
For block 1, the reasoning is the same, except that the information set is smaller and the information bits are $\{u_i^{(2)}\}_{i \in \mathcal{I}_v^{(1)} \cap \mathcal{I}^{(2)}}$. The complexity per block, under successive cancellation decoding, is $\Theta(N \log_2 N)$.

Performance. The block error probability $P_{\mathrm{B}}^{(l)}$ for the l-th user ($l \in \{1, 2\}$) can be upper bounded by

$$\begin{aligned} P_{\mathrm{B}}^{(1)} &\leq k \sum_{i \in \mathcal{I}_v^{(1)}} Z(U_i^{(2)} \mid U_{1:i-1}^{(2)}, Y_{1:N}^{(1)}) + k \sum_{i \in \mathcal{I}^{(1)}} Z(U_i^{(1)} \mid U_{1:i-1}^{(1)}, Y_{1:N}^{(1)}) = O(2^{-N^\beta}), \\ P_{\mathrm{B}}^{(2)} &\leq \sum_{i \in \mathcal{I}^{(2)}} Z(U_i^{(2)} \mid U_{1:i-1}^{(2)}, Y_{1:N}^{(2)}) = O(2^{-N^\beta}), \end{aligned} \tag{4.48}$$

for any $\beta \in (0, 1/2)$.

4.6 Polar Codes for Binning Region

The following theorem provides our main result regarding the achievability of the binning region for DM-BCs with polar codes (compare with Theorem 4.2).

Theorem 4.7 (Polar Codes for Binning Region)**.** *Consider a two-user DM-BC $p_{Y_1,Y_2|X}$, where X denotes the input to the channel, taking values on an arbitrary set $\mathcal{X}$, and Y_1, Y_2 denote the outputs at the first and second receiver, respectively. Let V_1 and V_2 denote auxiliary binary random variables. Then, for any joint distribution p_{V_1,V_2}, for any deterministic function $\phi\colon \{0,1\}^2 \to \mathcal{X}$ such that $X = \phi(V_1, V_2)$, and for any rate pair (R_1, R_2) satisfying the constraints (4.2), there exists a sequence of polar codes with an increasing block length N that achieves this rate pair with encoding and decoding complexity $\Theta(N \log_2 N)$ and a block error probability decaying like $O(2^{-N^\beta})$ for any $\beta \in (0, 1/2)$.*

Problem Statement. Let $(V_1, V_2) \sim p_{V_1,V_2} = p_{V_1} p_{V_2|V_1}$, and let X be a deterministic function ϕ of (V_1, V_2). The aim is to transmit over the two-user DM-BC $p_{Y_1,Y_2|X}$ achieving the rate pair

$$(R_1, R_2) = (I(V_1; Y_1), I(V_2; Y_2) - I(V_1; V_2)), \tag{4.49}$$

assuming that $I(V_1; V_2) < I(V_2; Y_2)$. Consequently, by Proposition 4.2, we can achieve the whole region (4.2) and Theorem 4.7 is proved. Note that if polar coding achieves the rate pair (4.49) with complexity $\Theta(N \log_2 N)$ and a block error probability $O(2^{-N^\beta})$, then for any other rate pair in the region (4.2), there exists a sequence of polar codes with an increasing block length N whose complexity and block error probability have the same asymptotic scalings.

Design of the Scheme. Set $U_{1:N}^{(1)} = V_{1:N}^{(1)} G_N$ and $U_{1:N}^{(2)} = V_{1:N}^{(2)} G_N$. As in the case of the transmission over a binary memoryless channel with V_l in place of X and Y_l in place of Y ($l \in \{1, 2\}$), define the sets $\mathcal{H}_{V_l}$, $\mathcal{L}_{V_l}$, $\mathcal{H}_{V_l|Y_l}$, and $\mathcal{L}_{V_l|Y_l}$ for $l \in \{1, 2\}$, similarly to (4.26) (except that we replace U_2 with U_l and V with V_l). These sets satisfy

$$\begin{aligned}
\lim_{N\to\infty} \frac{1}{N} |\mathcal{H}_{V_l}| &= H(V_l), \\
\lim_{N\to\infty} \frac{1}{N} |\mathcal{L}_{V_l}| &= 1 - H(V_l), \\
\lim_{N\to\infty} \frac{1}{N} |\mathcal{H}_{V_l|Y_l}| &= H(V_l \mid Y_l), \\
\lim_{N\to\infty} \frac{1}{N} |\mathcal{L}_{V_l|Y_l}| &= 1 - H(V_l \mid Y_l).
\end{aligned} \tag{4.50}$$

By thinking of V_1 as a side information for V_2, we can further define the sets $\mathcal{H}_{V_2|V_1}$ and $\mathcal{L}_{V_2|V_1}$ that satisfy

$$\begin{aligned}
\lim_{N\to\infty} \frac{1}{N} |\mathcal{H}_{V_2|V_1}| &= H(V_2 \mid V_1), \\
\lim_{N\to\infty} \frac{1}{N} |\mathcal{L}_{V_2|V_1}| &= 1 - H(V_2 \mid V_1).
\end{aligned} \tag{4.51}$$

First, consider only the point-to-point communication problem between the transmitter and the first receiver. As discussed in Section 4.4.2, for this scenario, the correct choice is to place the information in those positions of $U_{1:N}^{(1)}$ indexed by the set $\mathcal{I}^{(1)} = \mathcal{H}_{V_1} \cap \mathcal{L}_{V_1|Y_1}$ that satisfies

$$\lim_{N\to\infty} \frac{1}{N} |\mathcal{I}^{(1)}| = I(V_1; Y_1). \tag{4.52}$$

For the point-to-point communication problem between the transmitter and the second receiver, we know from Section 4.4.2 that the information has to be placed in those positions of $U_{1:N}^{(2)}$ that are indexed by $\mathcal{H}_{V_2} \cap \mathcal{L}_{V_2|Y_2}$.

Let us get back to the broadcasting scenario and note that for binning, unlike superposition coding, the first user does not decode the message intended for the second user. Consider the following scheme. The first user adopts the point-to-point communication strategy: it ignores the existence of the second user, and it uses $\mathcal{I}^{(1)}$ as an information set. The frozen positions are divided into the two usual subsets $\mathcal{F}_{\mathrm{d}}^{(1)} = \mathcal{H}_{V_1}^{\mathrm{c}}$ and $\mathcal{F}_{\mathrm{r}}^{(1)} = \mathcal{H}_{V_1} \cap \mathcal{L}_{V_1|Y_1}^{\mathrm{c}}$ that contain positions such that, respectively, $U_i^{(1)}$ can or cannot be approximately inferred from $U_{1:i-1}^{(1)}$. Whereas, the second user does not ignore the existence of the first user by putting his information in $\mathcal{H}_{V_2} \cap \mathcal{L}_{V_2|Y_2}$. Indeed, V_1 and V_2 are, in general, correlated. Hence, the second user puts his information in $\mathcal{I}^{(2)} = \mathcal{H}_{V_2|V_1} \cap \mathcal{L}_{V_2|Y_2}$. If $i \in \mathcal{I}^{(2)}$, then, since $\mathcal{I}^{(2)} \subseteq \mathcal{H}_{V_2|V_1}$, the bit $U_i^{(2)}$ is approximately independent of $(U_{1:i-1}^{(2)}, V_{1:N}^{(1)})$. This implies that $U_i^{(2)}$ is suitable to contain information. Furthermore, since $i \in \mathcal{L}_{V_2|Y_2}$, the bit $U_i^{(2)}$ is approximately a deterministic function of $(U_{1:i-1}^{(2)}, Y_{1:N}^{(2)})$. This implies that it is also decodable given the channel output $Y_{1:N}^{(2)}$. The remaining positions need to be frozen and can be divided into the four subsets described below.

- For $i \in \mathcal{F}_{\mathrm{r}}^{(2)} = \mathcal{H}_{V_2|V_1} \cap \mathcal{L}_{V_2|Y_2}^{\mathrm{c}}$, $U_i^{(2)}$ is chosen uniformly at random, and this randomness is shared between the transmitter and the second receiver.
- For $i \in \mathcal{F}_{\mathrm{d}}^{(2)} = \mathcal{L}_{V_2}$, $U_i^{(2)}$ is approximately a deterministic function of $U_{1:i-1}^{(2)}$, hence its value can be deduced from the past.
- For $i \in \mathcal{F}_{\mathrm{out}}^{(2)} = \mathcal{H}_{V_2|V_1}^{\mathrm{c}} \cap \mathcal{L}_{V_2}^{\mathrm{c}} \cap \mathcal{L}_{V_2|Y_2}$, $U_i^{(2)}$ is approximately a deterministic function of $(U_{1:i-1}^{(2)}, V_{1:N}^{(1)})$, but it can be deduced also from the channel output $Y_{1:N}^{(2)}$.
- For $i \in \mathcal{F}_{\mathrm{cr}}^{(2)} = \mathcal{H}_{V_2|V_1}^{\mathrm{c}} \cap \mathcal{L}_{V_2}^{\mathrm{c}} \cap \mathcal{L}_{V_2|Y_2}^{\mathrm{c}} = \mathcal{H}_{V_2|V_1}^{\mathrm{c}} \cap \mathcal{L}_{V_2|Y_2}^{\mathrm{c}}$, $U_i^{(2)}$ is approximately a deterministic function of $(U_{1:i-1}^{(2)}, V_{1:N}^{(1)})$, but it cannot be deduced either from $U_{1:i-1}^{(2)}$ or from $Y_{1:N}^{(2)}$.

The positions belonging to the last set are critical, since, in order to decode them, the receiver needs to know $V_{1:N}^{(1)}$. Indeed, recall that the encoding operation is performed jointly by the two users, whereas the first and the second decoder act separately and cannot exchange any information. The situation is schematically represented in Figure 4.8.

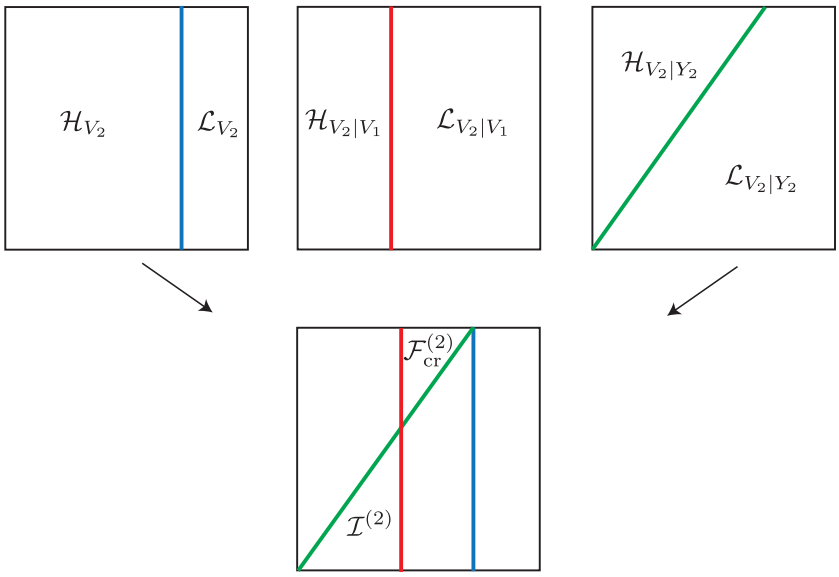


Figure 4.8 – Graphical representation of the sets associated with the second user for the binning scheme: $\mathcal{I}^{(2)}$ contains the information bits; $\mathcal{F}_{\text{cr}}^{(2)}$ contains the frozen positions that are critical in the sense that they cannot be inferred either from the past $U_{1:i-1}^{(2)}$ or from the channel output $Y_{1:N}^{(2)}$.

We start by reviewing the AGG scheme [100]. This scheme achieves the rate pair in (4.49), assuming that the degradation relation $p_{Y_2|V_2} \succ p_{V_1|V_2}$ holds. Note that, under this assumption, we have $\mathcal{L}_{V_2|V_1} \subseteq \mathcal{L}_{V_2|Y_2}$. Therefore, $\mathcal{F}_{\text{cr}}^{(2)} \subseteq \mathcal{L}_{V_2|V_1}^{\text{c}} \cap \mathcal{H}_{V_2|V_1}^{\text{c}}$. Since $|\mathcal{L}_{V_2|V_1}^{\text{c}} \cap \mathcal{H}_{V_2|V_1}^{\text{c}}| = o(N)$, it is assumed in [100] that the bits indexed by $\mathcal{L}_{V_2|V_1}^{\text{c}} \cap \mathcal{H}_{V_2|V_1}^{\text{c}}$ are "genie-given" from the encoder to the second decoder. The price to be paid for the transmission of these extra bits is asymptotically negligible. Consequently, the first user places his information in $\mathcal{I}^{(1)}$, the second user places his information in $\mathcal{I}^{(2)}$, and the bits in the positions belonging to $\mathcal{L}_{V_2|V_1}^{\text{c}} \cap \mathcal{H}_{V_2|V_1}^{\text{c}}$ are pre-communicated to the second receiver.

Our goal is to achieve the rate pair (4.49) without the degradation condition $p_{Y_2|V_2} \succ p_{V_1|V_2}$. As in the superposition coding scheme, the idea consists in transmitting k polar blocks and in repeating ("chaining") some bits from one block to the following block. To do so, let $\mathcal{R}$ be a subset of $\mathcal{I}^{(2)}$ such that $|\mathcal{R}| = |\mathcal{F}_{\text{cr}}^{(2)}|$. As usual, it does not matter what subset we pick. Since the second user cannot reconstruct the bits at the critical positions $\mathcal{F}_{\text{cr}}^{(2)}$, we use the set $\mathcal{R}$ to store the critical bits of the previous block. This construction is schematically represented in Figure 4.9.

Let us explain the scheme in detail. For block 1, we adopt the point-to-point communication strategy: the first user puts his information in $\mathcal{I}^{(1)}$, and the second user in $\mathcal{I}^{(2)}$. For block j ($j \in \{2, \cdots, k-1\}$), the first user places again his information in $\mathcal{I}^{(1)}$. The second user puts information in the positions indexed by $\mathcal{I}^{(2)} \setminus \mathcal{R}$ and repeats in $\mathcal{R}$ the bits that were contained in the set $\mathcal{F}_{\text{cr}}^{(2)}$ of block $j-1$. For block k, the second user does not change his strategy, putting information in $\mathcal{I}^{(2)} \setminus \mathcal{R}$ and repeating in $\mathcal{R}$ the bits that were contained in the set $\mathcal{F}_{\text{cr}}^{(2)}$ of block $k-1$. However, for block k, the first user does not convey any information and puts in $\mathcal{I}^{(1)}$ a fixed

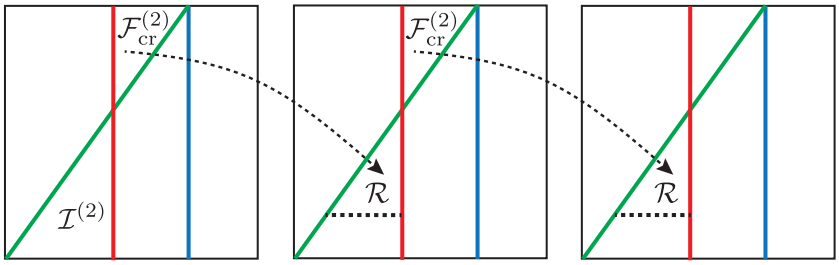


Figure 4.9 – Graphical representation of the chaining construction for the binning scheme with $k = 3$: the set $\mathcal{F}_{\mathrm{cr}}^{(2)}$ is repeated into the set $\mathcal{R}$ of the following block.

sequence that is shared between the encoder and both decoders. Indeed, in the last block, the positions indexed by $\mathcal{F}_{\mathrm{cr}}^{(2)}$ are not repeated anywhere. Consequently, the only way in which the second decoder can reconstruct the bits in $\mathcal{F}_{\mathrm{cr}}^{(2)}$ consists in knowing a priori the value of $V_{1:N}^{(1)}$.

Note that with this scheme, the second user has to decode "backwards", starting with block k and ending with block 1. In fact, for block k, the second user can compute $V_{1:N}^{(1)}$, hence the critical positions indexed by $\mathcal{F}_{\mathrm{cr}}^{(2)}$ are no longer a problem. Then, for block j ($j \in \{2, \cdots, k-1\}$), the second user knows the values of the bits in $\mathcal{F}_{\mathrm{cr}}^{(2)}$ from the decoding of the set $\mathcal{R}$ of block $j+1$.

Suppose now that the second user wants to decode "forward", i.e., starting with block 1 and ending with block k. Then, the set $\mathcal{R}$ is used to store the critical bits of the following block (instead of those ones of the previous block). In particular, for block k, we adopt the point-to-point communication strategy. For block j ($j \in \{k-1, \cdots, 2\}$), the first user places his information in $\mathcal{I}^{(1)}$, the second user places his information in the positions indexed by $\mathcal{I}^{(2)} \setminus \mathcal{R}$ and repeats in $\mathcal{R}$ the bits that were contained in the set $\mathcal{F}_{\mathrm{cr}}^{(2)}$ of block $j+1$. For block 1, the second user does not change his strategy, and the first user puts in $\mathcal{I}^{(1)}$ a shared fixed sequence. Note that in this case the encoding needs to be performed "backwards".

Encoding. Let us start from the first user. For block j ($j \in \{1, \cdots, k-1\}$):

- Let $\{u_i^{(1)}\}_{i \in \mathcal{I}^{(1)}}$ denote the information bits.
- Denote by $\{u_i^{(1)}\}_{i \in \mathcal{F}_{\mathrm{r}}^{(1)}}$ a particular realization of a random sequence that is shared between the transmitter and the first receiver.
- For $i \in \mathcal{F}_{\mathrm{d}}^{(1)}$, we can either use a "randomized rounding" rule, i.e.,

$$u_i^{(1)} = \begin{cases} 0, & \text{w.p. } \mathbb{P}_{U_i^{(1)}|U_{1:i-1}^{(1)}}(0 \mid u_{1:i-1}^{(1)}) \\ 1, & \text{w.p. } \mathbb{P}_{U_i^{(1)}|U_{1:i-1}^{(1)}}(1 \mid u_{1:i-1}^{(1)}) \end{cases} \tag{4.53}$$

  or an "argmax" rule, i.e.,

$$u_i^{(1)} = \arg \max_{u \in \{0,1\}} \mathbb{P}_{U_i^{(1)}|U_{1:i-1}^{(1)}}(u \mid u_{1:i-1}^{(1)}). \tag{4.54}$$

In case the "randomized rounding" rule is employed, the random sequence in (4.53) is shared between the transmitter and the receiver.

For block k:

- The user conveys no information, and $\{u_i^{(1)}\}_{i\in\mathcal{I}^{(1)}}$ contains a fixed sequence known to the second decoder.
- The frozen bits are chosen according to the usual rules with the only difference that the sequence $\{u_i^{(1)}\}_{i\in\mathcal{F}_{\rm r}^{(1)}}$ is shared also with the second decoder.

The rate of communication of the first user is given by

$$R_1 = \left(\frac{k-1}{kN}\right)|\mathcal{I}^{(1)}| = \left(\frac{k-1}{k}\right) I(V_1;Y_1) + o(1), \tag{4.55}$$

where we use (4.52). From (4.55), we obtain that, by choosing a large value of k, R_1 approaches $I(V_1;Y_1)$. Then, the vector $v_{1:N}^{(1)} = u_{1:N}^{(1)} G_N$ is obtained.

Let us now move to the second user. For block 1:

- Let $\{u_i^{(2)}\}_{i\in\mathcal{I}^{(2)}}$ denote the information bits.
- Denote by $\{u_i^{(2)}\}_{i\in\mathcal{F}_{\rm r}^{(2)}}$ a particular realization of a random sequence that is shared between the transmitter and the second receiver.
- For $i \in \mathcal{F}_{\rm d}^{(2)}$, we can either use a "randomized rounding" rule, i.e.,

$$u_i^{(2)} = \begin{cases} 0, & \text{w.p. } \mathbb{P}_{U_i^{(2)}|U_{1:i-1}^{(2)}}(0 \mid u_{1:i-1}^{(2)}) \\ 1, & \text{w.p. } \mathbb{P}_{U_i^{(2)}|U_{1:i-1}^{(2)}}(1 \mid u_{1:i-1}^{(2)}) \end{cases} \tag{4.56}$$

  or an "argmax" rule, i.e.,

$$u_i^{(2)} = \arg\max_{u\in\{0,1\}} \mathbb{P}_{U_i^{(2)}|U_{1:i-1}^{(2)}}(u \mid u_{1:i-1}^{(2)}). \tag{4.57}$$

  In case the "randomized rounding" rule is employed, the random sequence in (4.56) is shared between the transmitter and the receiver.

- For $i \in \mathcal{F}_{\rm out}^{(2)} \cup \mathcal{F}_{\rm cr}^{(2)}$, once again, we can either use a "randomized rounding" rule, i.e.,

$$u_i^{(2)} = \begin{cases} 0, & \text{w.p. } \mathbb{P}_{U_i^{(2)}|U_{1:i-1}^{(2)},V_{1:N}^{(1)}}(0 \mid u_{1:i-1}^{(2)}, v_{1:N}^{(1)}) \\ 1, & \text{w.p. } \mathbb{P}_{U_i^{(2)}|U_{1:i-1}^{(2)},V_{1:N}^{(1)}}(1 \mid u_{1:i-1}^{(2)}, v_{1:N}^{(1)}) \end{cases} \tag{4.58}$$

  or an "argmax" rule, i.e.,

$$u_i^{(2)} = \arg\max_{u\in\{0,1\}} \mathbb{P}_{U_i^{(2)}|U_{1:i-1}^{(2)},V_{1:N}^{(1)}}(u \mid u_{1:i-1}^{(2)}, v_{1:N}^{(1)}). \tag{4.59}$$

  Observe that the encoder has an access to $v_{1:N}^{(1)}$, hence it can compute the probabilities above. In case the "randomized rounding" rule is employed, the realization of the random sequence in (4.58) is shared between the transmitter and the receiver.

For block j ($j \in \{2, \cdots, k\}$):

- Let $\{u_i^{(2)}\}_{i \in \mathcal{I}^{(2)} \setminus \mathcal{R}}$ denote the information bits.
- The sequence $\{u_i^{(2)}\}_{i \in \mathcal{R}}$ contains a copy of the sequence $\{u_i^{(2)}\}_{i \in \mathcal{F}_{\mathrm{cr}}^{(2)}}$ of block $j-1$.
- The frozen bits are chosen as in block 1.

In order to compute the rate achievable by the second user, first observe that

$$\begin{aligned}
\frac{1}{N}\left(|\mathcal{I}^{(2)}| - |\mathcal{R}|\right) &\overset{\text{(a)}}{=} \frac{1}{N}\left(|\mathcal{H}_{V_2|V_1} \cap \mathcal{L}_{V_2|Y_2}| - |\mathcal{H}^{\mathrm{c}}_{V_2|V_1} \cap \mathcal{L}^{\mathrm{c}}_{V_2} \cap \mathcal{L}^{\mathrm{c}}_{V_2|Y_2}|\right) \\
&\overset{\text{(b)}}{=} \frac{1}{N}\Big(|(\mathcal{H}_{V_2} \cap \mathcal{L}_{V_2|Y_2}) \setminus (\mathcal{H}_{V_2} \cap \mathcal{H}^{\mathrm{c}}_{V_2|V_1})| \\
&\qquad - |(\mathcal{H}_{V_2} \cap \mathcal{H}^{\mathrm{c}}_{V_2|V_1}) \setminus (\mathcal{H}_{V_2} \cap \mathcal{L}_{V_2|Y_2})|\Big) + o(1) \\
&\overset{\text{(c)}}{=} \frac{1}{N}\left(|\mathcal{H}_{V_2} \cap \mathcal{L}_{V_2|Y_2}| - |\mathcal{H}_{V_2} \cap \mathcal{H}^{\mathrm{c}}_{V_2|V_1}|\right) + o(1) \\
&\overset{\text{(d)}}{=} \frac{1}{N}\left(|\mathcal{H}_{V_2} \cap \mathcal{L}_{V_2|Y_2}| - |\mathcal{H}_{V_2} \cap \mathcal{L}_{V_2|V_1}|\right) + o(1) \\
&\overset{\text{(e)}}{=} \frac{1}{N}\left(|\mathcal{L}_{V_2|Y_2} \setminus \mathcal{L}_{V_2}| - |\mathcal{L}_{V_2|V_1} \setminus \mathcal{L}_{V_2}|\right) + o(1) \\
&\overset{\text{(f)}}{=} I(V_2;Y_2) - I(V_1;V_2) + o(1),
\end{aligned} \tag{4.60}$$

where equality (a) holds since $|\mathcal{R}| = |\mathcal{F}_{\mathrm{cr}}^{(2)}|$; equality (b) follows from $\mathcal{H}_{V_2|V_1} \subseteq \mathcal{H}_{V_2}$ and $|[N] \setminus (\mathcal{H}_{V_2} \cup \mathcal{L}_{V_2})| = o(N)$; equality (c) follows from the identity in (4.34) for arbitrary finite sets; equality (d) holds since $|[N] \setminus (\mathcal{H}_{V_2|V_1} \cup \mathcal{L}_{V_2|V_1})| = o(N)$; equality (e) holds since $|[N] \setminus (\mathcal{H}_{V_2} \cup \mathcal{L}_{V_2})| = o(N)$; and equality (f) follows from the second and fourth equalities in (4.50), as well as from the second equality in (4.51). Consequently,

$$R_2 = \frac{1}{Nk}|\mathcal{R}| + I(V_2;Y_2) - I(V_1;V_2) + o(1). \tag{4.61}$$

Hence, as k tends to infinity, R_2 approaches the required rate $I(V_2;Y_2) - I(V_1;V_2)$. Then, the vector $v_{1:N}^{(2)} = u_{1:N}^{(2)} G_N$ is obtained. Finally, the vector $x_{1:N} = \phi(v_{1:N}^{(1)}, v_{1:N}^{(2)})$ is transmitted over the channel. The encoding complexity per block is $\Theta(N \log_2 N)$.

Decoding. Let us start from the first user that reconstructs $u_{1:N}^{(1)}$ from the channel output $y_{1:N}^{(1)}$. For each block, the decision rule is given by

$$\hat{u}_i^{(1)} = \begin{cases} u_i^{(1)}, & \text{if } i \in \mathcal{F}_{\mathrm{r}}^{(1)} \\ \arg\max\limits_{u \in \{0,1\}} \mathbb{P}_{U_i^{(1)}|U_{1:i-1}^{(1)}}(u \mid u_{1:i-1}^{(1)}), & \text{if } i \in \mathcal{F}_{\mathrm{d}}^{(1)} \\ \arg\max\limits_{u \in \{0,1\}} \mathbb{P}_{U_i^{(1)}|U_{1:i-1}^{(1)},Y_{1:N}^{(1)}}(u \mid u_{1:i-1}^{(1)}, y_{1:N}^{(1)}), & \text{if } i \in \mathcal{I}^{(1)} \end{cases} \tag{4.62}$$

The second user reconstructs $u_{1:N}^{(2)}$ from the channel output $y_{1:N}^{(2)}$. As explained before, the decoding goes "backwards", starting from block k and ending with block

1. For block k, the second decoder knows $v_{1:N}^{(1)}$. Hence, the decision rule is given by

$$\hat{u}_i^{(2)} = \begin{cases} u_i^{(2)}, & \text{if } i \in \mathcal{F}_{\mathrm{r}}^{(2)} \\ \arg\max\limits_{u\in\{0,1\}} \mathbb{P}_{U_i^{(2)}|U_{1:i-1}^{(2)}}(u \mid u_{1:i-1}^{(2)}), & \text{if } i \in \mathcal{F}_{\mathrm{d}}^{(2)} \\ \arg\max\limits_{u\in\{0,1\}} \mathbb{P}_{U_i^{(2)}|U_{1:i-1}^{(2)},V_{1:N}^{(1)}}(u \mid u_{1:i-1}^{(2)}, v_{1:N}^{(1)}), & \text{if } i \in \mathcal{F}_{\mathrm{out}}^{(2)} \cup \mathcal{F}_{\mathrm{cr}}^{(2)} \\ \arg\max\limits_{u\in\{0,1\}} \mathbb{P}_{U_i^{(2)}|U_{1:i-1}^{(2)},Y_{1:N}^{(2)}}(u \mid u_{1:i-1}^{(2)}, y_{1:N}^{(2)}), & \text{if } i \in \mathcal{I}^{(2)} \end{cases} \tag{4.63}$$

In (4.62) and (4.63), we assume that the "argmax" rule is employed to encode the positions in $\mathcal{F}_{\mathrm{d}}^{(1)}$, $\mathcal{F}_{\mathrm{d}}^{(2)}$ and $\mathcal{F}_{\mathrm{out}}^{(2)} \cup \mathcal{F}_{\mathrm{cr}}^{(2)}$. Clearly, correct recovery is also possible under the "randomized rounding" rule.
For block j ($j \in \{2, \cdots, k\}$), the decision rule is the same as (4.63) for $i \notin \mathcal{F}_{\mathrm{out}}^{(2)} \cup \mathcal{F}_{\mathrm{cr}}^{(2)}$. Indeed, $\{u_i^{(2)}\}_{i\in\mathcal{F}_{\mathrm{cr}}^{(2)}}$ of block j can be deduced from $\{u_i^{(2)}\}_{i\in\mathcal{R}}$ of block $j+1$, and, for $i \in \mathcal{F}_{\mathrm{out}}^{(2)}$, we set

$$\hat{u}_i^{(2)} = \arg\max_{u\in\{0,1\}} \mathbb{P}_{U_i^{(2)}|U_{1:i-1}^{(2)},Y_{1:N}^{(2)}}(u \mid u_{1:i-1}^{(2)}, y_{1:N}^{(2)}).$$

The complexity per block, under successive cancellation decoding, is $\Theta(N \log_2 N)$.

Performance. The block error probability $P_{\mathrm{B}}^{(l)}$ for the l-th user ($l \in \{1,2\}$) can be upper bounded by

$$\begin{aligned} P_{\mathrm{B}}^{(1)} &\le k \sum_{i\in\mathcal{I}^{(1)}} Z(U_i^{(1)} \mid U_{1:i-1}^{(1)}, Y_{1:N}^{(1)}) = O(2^{-N^\beta}), \\ P_{\mathrm{B}}^{(2)} &\le k \sum_{i\in\mathcal{L}_{V_2|Y_2}} Z(U_i^{(2)} \mid U_{1:i-1}^{(2)}, Y_{1:N}^{(2)}) = O(2^{-N^\beta}), \end{aligned} \tag{4.64}$$

for any $\beta \in (0, 1/2)$.

4.7 Polar Codes for Marton's Region

4.7.1 Only Private Messages

Consider first the case where only private messages are available. The following theorem provides our main result regarding the achievability with polar codes of Marton's region that forms the tightest inner bound known to date for a two-user DM-BC without common information (compare with Theorem 4.3).

Theorem 4.8 (Polar Codes for Marton's Region)**.** *Consider a two-user DM-BC $p_{Y_1,Y_2|X}$, where X denotes the input to the channel, taking values on an arbitrary set $\mathcal{X}$, and Y_1, Y_2 denote the outputs at the first and second receiver, respectively. Let V, V_1, and V_2 denote auxiliary binary random variables. Then, for any joint distribution p_{V,V_1,V_2}, for any deterministic function $\phi : \{0,1\}^3 \to \mathcal{X}$ such that $X = \phi(V, V_1, V_2)$, and for any rate pair (R_1, R_2) satisfying the constraints* (4.3)*, there exists a sequence of polar codes with an increasing block length N that achieves this rate pair with encoding and decoding complexity $\Theta(N \log_2 N)$ and a block error probability decaying like $O(2^{-N^\beta})$ for any $\beta \in (0, 1/2)$.*

The proposed coding scheme is a combination of the techniques described in detail in Sections 4.5 and 4.6, and it is outlined below.

Problem Statement. Let $(V, V_1, V_2) \sim p_V\, p_{V_2|V}\, p_{V_1|V_2V}$, and let X be a deterministic function of (V, V_1, V_2), i.e., $X = \phi(V, V_1, V_2)$. Consider the two-user DM-BC $p_{Y_1,Y_2|X}$ such that $I(V;Y_1) \leq I(V;Y_2)$. The aim is to achieve the rate pair

$$(R_1, R_2) = (I(V, V_1; Y_1) - I(V_1; V_2 \mid V) - I(V; Y_2), I(V, V_2; Y_2)). \tag{4.65}$$

Once we have accomplished this, we will see that a slight modification of this scheme enables us to achieve, in addition, the rate pair

$$(R_1, R_2) = (I(V, V_1; Y_1), I(V_2; Y_2 \mid V) - I(V_1; V_2 \mid V)). \tag{4.66}$$

Therefore, by Proposition 4.3, we can achieve the whole rate region in (4.3) by using polar codes. Note that if polar coding achieves the rate pairs (4.65) and (4.66) with complexity $\Theta(N \log_2 N)$ and a block error probability $O(2^{-N^\beta})$, then for any other rate pair in the region (4.3), there exists a sequence of polar codes with an increasing block length N whose complexity and block error probability have the same asymptotic scalings.

Sketch of the Scheme. Set $U_{1:N}^{(0)} = V_{1:N}G_N$, $U_{1:N}^{(1)} = V_{1:N}^{(1)}G_N$, and $U_{1:N}^{(2)} = V_{1:N}^{(2)}G_N$. Then, the idea is that $U_{1:N}^{(1)}$ carries the message of the first user, whereas $U_{1:N}^{(0)}$ and $U_{1:N}^{(2)}$ carry the message of the second user. On the one hand, the first user will decode both his message, namely, $U_{1:N}^{(1)}$, and a part of the message of the second user, namely, $U_{1:N}^{(0)}$. On the other hand, the second user will be able to decode only his message, namely, $U_{1:N}^{(0)}$ and $U_{1:N}^{(2)}$. The random variable V comes from the superposition coding scheme, because $U_{1:N}^{(0)}$ is decodable by both users, but carries information meant only for one of them. The random variables V_1 and V_2 come from the binning scheme, since the first user decodes $U_{1:N}^{(1)}$ and the second user decodes $U_{1:N}^{(2)}$, i.e., each user decodes only his own information.

Let the sets $\mathcal{H}_V$, $\mathcal{L}_V$, $\mathcal{H}_{V|Y_l}$, and $\mathcal{L}_{V|Y_l}$ for $l \in \{1, 2\}$ be defined as in (4.26), where these subsets of $[N]$ satisfy (4.27). In analogy to Sections 4.5 and 4.6 let us also define the following sets ($l \in \{1, 2\}$):

$$\begin{aligned}
\mathcal{H}_{V_l|V} &= \{i \in [N] \colon Z(U_i^{(l)} \mid U_{1:i-1}^{(l)}, U_{1:N}^{(0)}) \geq 1 - \delta_N\},\\
\mathcal{L}_{V_l|V} &= \{i \in [N] \colon Z(U_i^{(l)} \mid U_{1:i-1}^{(l)}, U_{1:N}^{(0)}) \leq \delta_N\},\\
\mathcal{H}_{V_l|V,Y_l} &= \{i \in [N] \colon Z(U_i^{(l)} \mid U_{1:i-1}^{(l)}, U_{1:N}^{(0)}, Y_{1:N}^{(l)}) \geq 1 - \delta_N\},\\
\mathcal{L}_{V_l|V,Y_l} &= \{i \in [N] \colon Z(U_i^{(l)} \mid U_{1:i-1}^{(l)}, U_{1:N}^{(0)}, Y_{1:N}^{(l)}) \leq \delta_N\},\\
\mathcal{H}_{V_1|V,V_2} &= \{i \in [N] \colon Z(U_i^{(l)} \mid U_{1:i-1}^{(l)}, U_{1:N}^{(0)}, U_{1:N}^{(2)}) \geq 1 - \delta_N\},\\
\mathcal{L}_{V_1|V,V_2} &= \{i \in [N] \colon Z(U_i^{(l)} \mid U_{1:i-1}^{(l)}, U_{1:N}^{(0)}, U_{1:N}^{(2)}) \leq \delta_N\},
\end{aligned} \tag{4.67}$$

which satisfy

$$\begin{aligned}
&\lim_{N\to\infty}\frac{1}{N}\left|\mathcal{H}_{V_l|V}\right| = H(V_l \mid V),\\
&\lim_{N\to\infty}\frac{1}{N}\left|\mathcal{L}_{V_l|V}\right| = 1 - H(V_l \mid V),\\
&\lim_{N\to\infty}\frac{1}{N}\left|\mathcal{H}_{V_l|V,Y_l}\right| = H(V_l \mid V, Y_l),\\
&\lim_{N\to\infty}\frac{1}{N}\left|\mathcal{L}_{V_l|V,Y_l}\right| = 1 - H(V_l \mid V, Y_l),\\
&\lim_{N\to\infty}\frac{1}{N}\left|\mathcal{H}_{V_1|V,V_2}\right| = H(V_1 \mid V, V_2),\\
&\lim_{N\to\infty}\frac{1}{N}\left|\mathcal{L}_{V_1|V,V_2}\right| = 1 - H(V_1 \mid V, V_2).
\end{aligned} \tag{4.68}$$

First, consider the subsets of positions of $U_{1:N}^{(0)}$. The set $\mathcal{I}_{\text{sup}}^{(2)} = \mathcal{H}_V \cap \mathcal{L}_{V|Y_2}$ contains the positions decodable by the second user, and the set $\mathcal{I}_v^{(1)} = \mathcal{H}_V \cap \mathcal{L}_{V|Y_1}$ contains the positions decodable by the first user. Recall that $U_{1:N}^{(0)}$ needs to be decoded by both users, but contains information only for the second user.

Second, consider the subsets of positions of $U_{1:N}^{(2)}$. The set $\mathcal{I}_{\text{bin}}^{(2)} = \mathcal{H}_{V_2|V} \cap \mathcal{L}_{V_2|V,Y_2}$ contains the positions decodable by the second user. Recall that $U_{1:N}^{(2)}$ needs to be decoded only by the second user, and it contains part of his message.

Third, consider the subsets of positions of $U_{1:N}^{(1)}$. The set $\mathcal{I}^{(1)} = \mathcal{H}_{V_1|V,V_2} \cap \mathcal{L}_{V_1|V,Y_1}$ contains the positions decodable by the first user. Recall that $U_{1:N}^{(1)}$ needs to be decoded by the first user, and it contains only his message. However, the first user cannot decode $U_{1:N}^{(2)}$, hence it cannot infer $V_{1:N}^{(2)}$. Consequently, the positions in the set $\mathcal{F}_{\text{cr}}^{(1)} = \mathcal{H}_{V_1|V,V_2}^c \cap \mathcal{L}_{V_1|V}^c \cap \mathcal{L}_{V_1|V,Y_1}^c$ are critical. Indeed, for $i \in \mathcal{F}_{\text{cr}}^{(1)}$, the bit $U_i^{(1)}$ is approximately a deterministic function of $(U_{1:i-1}^{(1)}, U_{1:N}^{(0)}, U_{1:N}^{(2)})$, but it cannot be deduced from $(U_{1:i-1}^{(1)}, U_{1:N}^{(0)}, Y_{1:N}^{(1)})$.

In order to achieve the rate pair (4.65), we consider the transmission of k polar blocks and use three different "chaining" constructions. The first and the second chaining come from superposition coding, and the last one comes from binning.

First, define $\mathcal{D}^{(2)} = \mathcal{I}_{\text{sup}}^{(2)} \setminus \mathcal{I}_v^{(1)}$ and $\mathcal{D}^{(1)} = \mathcal{I}_v^{(1)} \setminus \mathcal{I}_{\text{sup}}^{(2)}$, as in (4.32) and (4.33), respectively. The former set contains the positions of $U_{1:N}^{(0)}$ that are decodable by the second user but not by the first, whereas the latter contains the positions of $U_{1:N}^{(0)}$ that are decodable by the first user but not by the second. Let $\mathcal{R}_{\text{sup}}$ be a subset of $\mathcal{D}^{(2)}$ such that $|\mathcal{R}_{\text{sup}}| = |\mathcal{D}^{(1)}|$. In block 1, fill $\mathcal{D}^{(1)}$ with information for the second user, and set the bits indexed by $\mathcal{D}^{(2)}$ to a fixed known sequence. In block j ($j \in \{2, \cdots, k-1\}$), fill $\mathcal{D}^{(1)}$ again with information for the second user, and repeat the bits contained in the set $\mathcal{D}^{(1)}$ of block $j-1$ into the positions indexed by $\mathcal{R}_{\text{sup}}$ of block j. In the final block k, put a known sequence in the positions indexed by $\mathcal{D}^{(1)}$, and repeat in the positions indexed by $\mathcal{R}_{\text{sup}}$ the bits contained in the set $\mathcal{D}^{(1)}$ of block $k-1$. In all the blocks, fill $\mathcal{I}_v^{(1)} \cap \mathcal{I}_{\text{sup}}^{(2)}$ with information for the second user. In this way, both users will be able to decode a fraction of the bits of $U_{1:N}^{(0)}$ that is roughly equal to $I(V; Y_1)$. The bits in these positions contain information for the second user.

Second, define $\mathcal{B}^{(2)} = \mathcal{D}^{(2)} \setminus \mathcal{R}_{\text{sup}}$, and let $\mathcal{B}^{(1)}$ be a subset of $\mathcal{I}^{(1)}$ such that $|\mathcal{B}^{(1)}| = |\mathcal{B}^{(2)}|$. Note that $\mathcal{B}^{(2)}$ contains positions of $U_{1:N}^{(0)}$, and $\mathcal{B}^{(1)}$ contains positions of $U_{1:N}^{(1)}$. For block j ($j \in \{2, \cdots, k\}$), we fill $\mathcal{B}^{(2)}$ with information for the second user, and we repeat these bits into the positions indexed by $\mathcal{B}^{(1)}$ of block $j-1$. In this way, both users will be able to decode a fraction of the bits of $U_{1:N}^{(0)}$ that is roughly equal to $I(V;Y_2)$ (recall that $I(V;Y_1) \leq I(V;Y_2)$). Again, the bits in these positions contain information for the second user.

Third, let $\mathcal{R}_{\text{bin}}$ be a subset of $\mathcal{I}^{(1)}$ such that $|\mathcal{R}_{\text{bin}}| = |\mathcal{F}_{\text{cr}}^{(1)}|$. Since the first user cannot reconstruct the bits at the critical positions $\mathcal{F}_{\text{cr}}^{(1)}$, we use the set $\mathcal{R}_{\text{bin}}$ to store the critical bits of the following block. For block k, the first user places all his information in $\mathcal{I}^{(1)}$. For block j ($j \in \{1, \cdots, k-1\}$), the first user places all his information in $\mathcal{I}^{(1)} \setminus (\mathcal{R}_{\text{bin}} \cup \mathcal{B}^{(1)})$, repeats in $\mathcal{R}_{\text{bin}}$ the bits in $\mathcal{F}_{\text{cr}}^{(1)}$ for block $j+1$, and repeats in $\mathcal{B}^{(1)}$ the bits in $\mathcal{B}^{(2)}$ for block $j+1$. The second user puts part of his information in $\mathcal{I}_{\text{bin}}^{(2)}$ (a subset of the positions of $U_{1:N}^{(2)}$) for all the blocks except for the first, in which $\mathcal{I}_{\text{bin}}^{(2)}$ contains a fixed sequence shared between the encoder and both decoders. Indeed, for block 1, the positions indexed by $\mathcal{F}_{\text{cr}}^{(1)}$ are not repeated anywhere, and the only way in which the first decoder can reconstruct those bits consists in knowing a-priori the value of $V_{1:N}^{(2)}$. The situation is schematically represented in Figures 4.10 and 4.11.

The encoding of $U_{1:N}^{(0)}$ is performed "forward", i.e., from block 1 to block k; the encoding of $U_{1:N}^{(1)}$ is performed "backwards", i.e., from block k to block 1; the encoding of $U_{1:N}^{(2)}$ can be performed in any order. The first user decodes $U_{1:N}^{(0)}$ and $U_{1:N}^{(1)}$ "forward"; the second user decodes $U_{1:N}^{(0)}$ "backwards" and can decode $U_{1:N}^{(2)}$ in any order.

With this polar coding scheme, by letting k tend to infinity, the first user decodes a fraction of the positions of $U_{1:N}^{(1)}$ containing his own message, given by

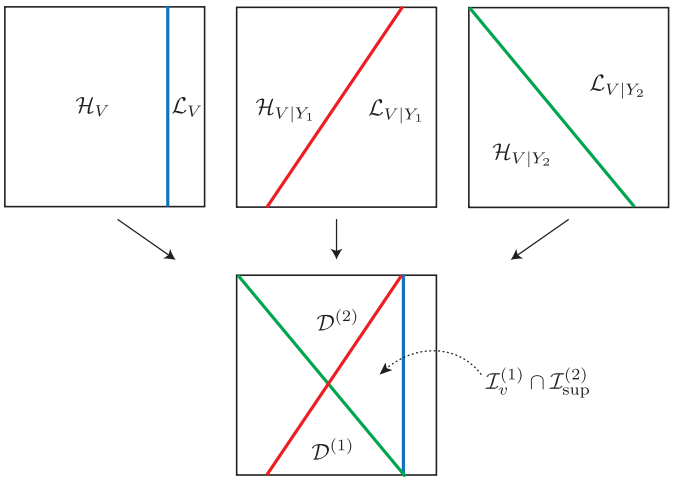
$$\begin{aligned} R_1 &= \frac{1}{N}(|\mathcal{I}^{(1)}| - |\mathcal{R}_{\text{bin}}| - |\mathcal{B}^{(1)}|) && (4.69)\\ &= I(V_1;Y_1 \mid V) - I(V_1;V_2 \mid V) - (I(V;Y_2) - I(V;Y_1)) \\ &= I(V,V_1;Y_1) - I(V_1;V_2 \mid V) - I(V;Y_2). && (4.70) \end{aligned}$$

The information for the second user is spread between the positions of $U_{1:N}^{(0)}$ and the positions of $U_{1:N}^{(2)}$ for a total rate that, as k tends to infinity, is given by

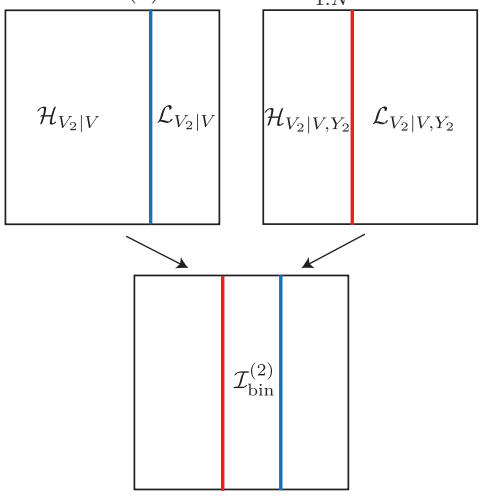
$$\begin{aligned} R_2 &= \frac{1}{N}(|\mathcal{I}_{\text{sup}}^{(2)}| + |\mathcal{I}_{\text{bin}}^{(2)}|) && (4.71)\\ &= I(V;Y_2) + I(V_2;Y_2 \mid V) && (4.72)\\ &= I(V,V_2;Y_2). && (4.73) \end{aligned}$$

It is possible to achieve the rate pair (4.66) with a scheme similar to the one described above by swapping the roles of the two users. Since $I(V;Y_1) \leq I(V;Y_2)$, only the first and the third chaining constructions are required. Indeed, the set that has the role of $\mathcal{B}^{(2)}$ is empty in this scenario.

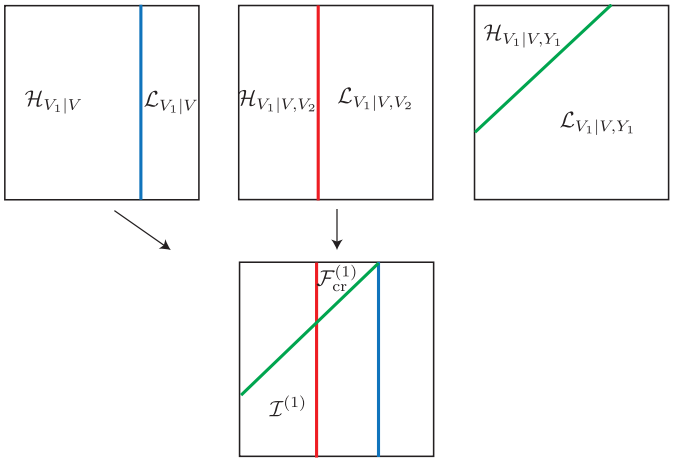
As our schemes consist in the repetition of polar blocks, the encoding and decoding complexity per block is $\Theta(N \log_2 N)$, and the block error probability decays like $O(2^{-N^\beta})$ for any $\beta \in (0, 1/2)$.



(a) Subsets of $U_{1:N}^{(0)}$.



(b) Subsets of $U_{1:N}^{(2)}$.



(c) Subsets of $U_{1:N}^{(1)}$.

Figure 4.10 – Graphical representation of the sets associated with the three auxiliary random variables in the scheme that achieves Marton's region with only private messages (4.3).

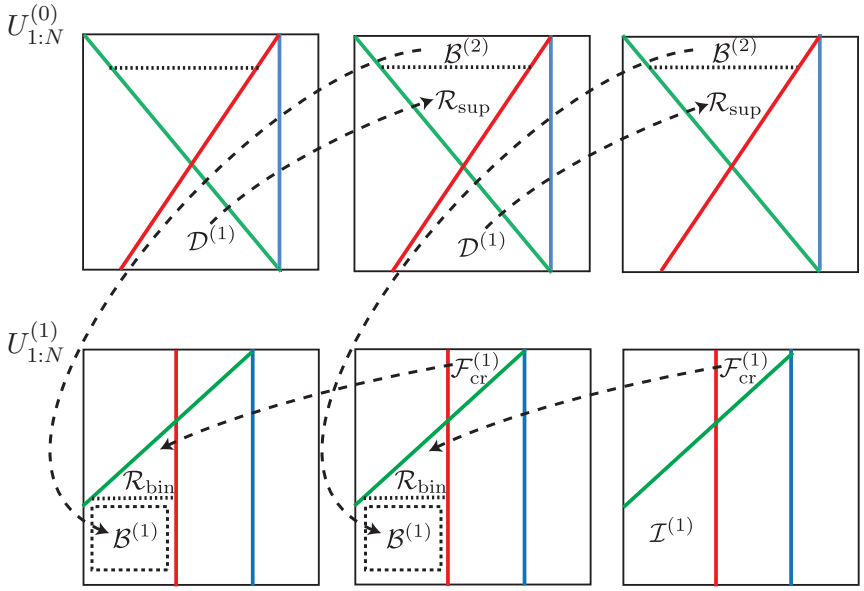


Figure 4.11 – Graphical representation of the chaining constructions for Marton's region with only private messages with $k = 3$: the set $\mathcal{D}^{(1)}$ is repeated into the set $\mathcal{R}_{\text{sup}}$ of the following block; the set $\mathcal{B}^{(2)}$ is repeated into the set $\mathcal{B}^{(1)}$ of the previous block; the set $\mathcal{F}_{\text{cr}}^{(1)}$ is repeated into the set $\mathcal{R}_{\text{bin}}$ of the previous block.

4.7.2 Private and Common Messages: MGP Region

Finally, consider the case of a two-user DM-BC with both common and private messages. Our most general result consists in the construction of polar codes that achieve the MGP region (4.4).

Theorem 4.9 (Polar Codes for MGP Region)**.** *Consider a two-user DM-BC $p_{Y_1,Y_2|X}$, where X denotes the input to the channel, taking values on an arbitrary set $\mathcal{X}$, and Y_1, Y_2 denote the outputs at the first and second receiver, respectively. Let R_0, R_1, and R_2 designate the rates of the common message and of the private messages of the two users, respectively. Let V, V_1, and V_2 denote auxiliary binary random variables. Then, for any joint distribution p_{V,V_1,V_2}, for any deterministic function $\phi\colon \{0,1\}^3 \to \mathcal{X}$ such that $X = \phi(V, V_1, V_2)$, and for any rate triple (R_0, R_1, R_2) satisfying the constraints (4.4), there exists a sequence of polar codes with an increasing block length N that achieves this rate triple with encoding and decoding complexity $\Theta(N \log_2 N)$ and a block error probability decaying like $O(2^{-N^\beta})$ for any $\beta \in (0, 1/2)$.*

The polar coding scheme follows the ideas outlined in Section 4.7.1. Recall that $U_{1:N}^{(0)}$ is decoded by both users. Then, we put the common information in the positions of $U_{1:N}^{(0)}$ that previously contained private information meant only for one of the users. The common rate is clearly upper bounded by $\min\{I(V;Y_1), I(V;Y_2)\}$. The remaining four inequalities of (4.4) are equivalent to the conditions in (4.3) with the only difference that a portion R_0 of the private information for one of the users has been converted into common information. This suffices to achieve the required rate region.

How to Achieve the Capacity of Asymmetric Channels

5

Sii sempre più o meno specifico.

Always be somehow specific.

In the previous chapter, we have presented low-complexity polar coding schemes for the broadcast channel. Our constructions are based on two polar "primitives", namely, lossless compression and transmission over an asymmetric channel. In this chapter[1], we focus specifically on asymmetric channels, and we survey three general capacity-achieving paradigms that provide a variety of provably optimal coding solutions.

In Section 5.1 we review the existing literature and summarize our main contributions. In Section 5.2 we discuss two coding "primitives", specifically, how to achieve the symmetric capacity of an asymmetric channel and how to perform error correction using biased codewords. The solutions to these problems will be later used as basic building blocks to devise coding schemes for asymmetric channels. Note that this modular strategy is very much similar to the one presented in the previous chapter for broadcast channels. Then, we describe in the next three consecutive sections the coding paradigms for achieving the capacity of an arbitrary (hence, possibly asymmetric) discrete memoryless channel (DMC): Gallager's mapping in Section 5.3, the integrated scheme in Section 5.4, and the chaining construction in Section 5.5. In Section 5.6, we provide a comparison between these three different approaches. The metrics taken into account are the error probability, the rate penalty, the computational complexity, the universality, the delay, and the use of common randomness. We defer the proofs of some results to the appendix in Section 5.7.

[1] The material of this chapter is based on joint work with S. H. Hassani and R. Urbanke [166,167].

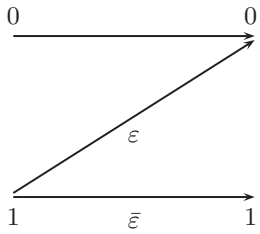


Figure 5.1 – Schematic representation of the Z-channel with parameter ε.

5.1 Related Work and Main Results

The simplest example of *asymmetric* DMC is the Z-channel that is schematically represented in Figure 5.1: the input symbol 0 is left untouched by the channel, whereas the input symbol 1 is flipped with probability ε. The basic problem that we face when transmitting over this channel is that (proper) linear codes impose a uniform input distribution, whereas the capacity-achieving input distribution for an asymmetric channel is, in general, non-uniform. Indeed, for the case of the Z-channel, the capacity-achieving distribution assigns to the symbol 1 a probability $\varepsilon^{\varepsilon/\bar{\varepsilon}} \cdot (1+\bar{\varepsilon}\varepsilon^{\varepsilon/\bar{\varepsilon}})^{-1}$, where we set $\bar{\varepsilon} = 1-\varepsilon$, see formula (5.10) of [44]. This mismatch in the input distribution bounds the achievable transmission rate away from capacity.

It is worth pointing out that, at least for binary inputs, the optimal distribution is not too far from the uniform one, in the sense that the capacity-achieving input distribution always has a marginal in the interval $(1/e, 1-1/e)$ [168]. In addition, a fraction of at most $1 - \frac{1}{2}e\ln(2) \approx 0.058$ of capacity is lost if we use the uniform input distribution instead of the optimal input distribution [168]. This result was later strengthened in [169], where the Z-channel is proved to be extremal in this sense. As for channels with more than 2 inputs, the upper bound $1-1/e$ to the range of the capacity-achieving distribution still holds [170], but the lower bound $1/e$ is false.

Given that the loss incurred by using a uniform input distribution is relatively modest, why do we care about the problem of achieving the full capacity of asymmetric channels? First of all, it is an interesting theoretical question. Second, over time, all communication systems are increasingly optimized to take full advantage of their capabilities, and even small gains become significant.

The classic solution to the problem of coding over asymmetric channels goes back to Gallager and consists of concatenating a linear code with a non-linear mapper so that the input distribution becomes biased [120]. In [40], McEliece described how this can be done successfully with iterative codes. We refer to this approach as *Gallager's mapping* and we discuss how any capacity-achieving coding scheme can be used for this setting. In particular, by combining either polar codes or spatially coupled codes with suitable non-linear mappers, we can approach capacity arbitrarily closely. More specifically, we derive a scaling law that relates the gap to capacity to the mismatch in the actual input distribution and to the size of the channel alphabets.

More recently polar codes have been used to achieve the capacity of binary-input

asymmetric DMCs. In particular, in [165] the authors propose a solution that makes use of the concatenation of two polar codes: one of them is used to solve a source coding problem, in order to have codewords distributed according to the capacity-achieving input distribution; the other is used to solve a channel coding problem, in order to provide error correction. However, such a scheme requires polarization for both the inner and the outer codes, therefore the error probability scales roughly as $2^{-N^{1/4}}$, where N is the block length of the code. Thus, in order to obtain the same performance as standard polar codes, the square of their block length is required. A very simple and more efficient solution for the transmission over asymmetric channels is presented in [121] and we have reviewed this idea in Section 4.4.2. More specifically, in order to transmit over channels whose optimal input distribution is non-uniform, the polar indices are partitioned into three groups: some are used for information transmission; some are used to ensure that the input distribution is properly biased; and some carry random bits shared between the transmitter and the receiver. The purpose of this shared randomness is to facilitate the performance analysis. Indeed, as in the case of LDPC codes, the error probability is obtained by averaging over the randomness of the ensemble. In short, the methods in [121, 165] exploit the fact that polar codes are well suited not only for channel coding but also for lossless source coding [91, 92]. Clearly, this is not a prerogative only of polar codes as, for example, sparse graph codes have been successfully used for both channel coding and source coding purposes [171]. Motivated by this fact, we describe a scheme based on spatially coupled codes that achieves the capacity of asymmetric DMCs by solving both a source coding and a channel coding problem at the same time. As it will be explained in detail later, this last solution still does not have a formal proof. We refer to this approach as the *integrated scheme*, because we use one code for both source *and* channel coding.

This brings us to the third coding paradigm. By "chaining" together several codewords, we can decouple the problem of source coding (creating a biased codeword from unbiased bits) from the problem of channel coding (providing error correction). The idea is based on [122], where the authors refer to it as the *bootstrap* scheme. We prefer to use the name *chaining construction* that was introduced in [48], where a similar approach was used to design universal polar codes. The chaining construction is a general method and we have already used it to devise polar coding schemes for the broadcast channel in Chapter 4 of this thesis. Here, we show how to chain *any* suitable source coding solution with *any* suitable channel coding solution, in order to transmit over an asymmetric channel. We give explicit conditions on the source and the channel code so that the overall scheme is capacity-achieving, and we prove that both polar codes and spatially coupled codes satisfy these conditions.

In summary, this chapter surveys **three different paradigms to achieve the capacity of asymmetric channels** and, as such, it is of tutorial nature. Motivated by the recent advances in coding for symmetric channels, we show that it is now possible to construct efficient schemes also for the asymmetric case. As a result, we demonstrate that perhaps what was once considered as a difficult problem is in fact quite easy to solve with existing "primitives". The three paradigms presented are quite general and should be regarded as "meta-schemes" that can then be made more specific by using a certain class of codes (e.g., polar codes or spatially coupled codes) according to the particular scenario of interest. For this reason, the interest of this chapter is more in describing generic coding ideas rather than in presenting

formal proofs and providing all the details for each scheme. With the objective of highlighting the pros and cons of these coding approaches, we present them in a unified manner and provide a detailed comparison with a focus on several features crucial in applications, such as, error probability, rate penalty, computational complexity, universality, delay, and access to common randomness.

5.2 Two Coding Primitives

First of all, we establish the notation and review some known concepts. Part of this notation has been already introduced, but, for the sake of clarity, we define again all the symbols that we use in the rest of the chapter. Then, we consider two problems that will be regarded as useful primitives. In particular, in Section 5.2.2, we discuss how to achieve the symmetric capacity of an asymmetric binary-input DMC (B-DMC), and in Section 5.2.3 we describe how to transmit reliably a biased binary codeword.

5.2.1 Notation and Prerequisites

Throughout this chapter, we consider the transmission over a DMC $W : \mathcal{X} \to \mathcal{Y}$ with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$. If the channel is binary-input, we usually take $\mathcal{X} = \mathbb{F}_2 = \{0, 1\}$ and we say that X is a Bernoulli(α) random variable if $\mathbb{P}(X = 1) = \alpha$ for some $\alpha \in [0, 1]$. However, for the analysis of LDPC ensembles, it is convenient to consider the standard mapping $0 \longleftrightarrow 1$ and $1 \longleftrightarrow -1$. It will be clear from the context whether the input alphabet is $\{-1, 1\}$ or $\{0, 1\}$. The probability of the output being y given an input x is denoted by $W(y \mid x)$ and the probability of the input being x given an output y is denoted by $p_{X|Y}(x \mid y)$. We write $C(W)$ and $C_{\mathrm{s}}(W)$ to indicate the capacity and the symmetric capacity of W, respectively. Given the scalar components $X_i, \cdots, X_j$ and $X^{(i)}, \cdots, X^{(j)}$, we use $X_{i:j}$ as a shorthand for the row vector $(X_i, \cdots, X_j)$ and, similarly, $X^{i:j}$ as a shorthand for the row vector $(X^{(i)}, \cdots, X^{(j)})$ with $i \leq j$. The index set $\{1, \cdots, N\}$ is abbreviated as $[N]$ and, given a set $\mathcal{A} \subseteq [N]$, we denote by $\mathcal{A}^{\mathrm{c}}$ its complement. We denote by $\log_2$ and ln the logarithm in base 2 and base e, respectively. For any $x \in [0, 1]$, we define $\bar{x} = 1 - x$. The binary entropy function is given by $h_2(x) = -x \log_2 x - \bar{x} \log_2 \bar{x}$. When discussing sparse graph coding schemes, we denote the parity-check matrix and its transpose by P and P^T, respectively. We do not use H to denote the parity-check matrix, as it is done more frequently, because the symbol $H(\cdot)$ indicates the entropy of a random variable. When discussing polar coding schemes, we assume that the block length N is a power of 2, say $N = 2^n$ for $n \in \mathbb{N}$, and we denote by G_N the generator matrix defined in (1.13).

Let us recall some basic facts concerning B-DMCs. This part is telegraphic and the reader is referred to [44] for more details. First, consider a symmetric B-DMC with $\mathcal{X} = \{-1, 1\}$. Assume that X is transmitted, Y is the received observation, and $L(Y)$ the corresponding log-likelihood ratio, namely for any $y \in \mathcal{Y}$,

$$L(y) = \ln \frac{W(y \mid 1)}{W(y \mid -1)}. \tag{5.1}$$

Let us denote by a the density of $L(Y)$ assuming that $X = 1$ and let us call it an L-density.

We say that an L-density a is symmetric if

$$\mathsf{a}(y) = e^y \mathsf{a}(-y). \tag{5.2}$$

Since the log-likelihood ratio constitutes a sufficient statistic for decoding, two symmetric B-DMCs are equivalent if they have the same L-density. A meaningful choice for the representative of each equivalence class is $W(y \mid 1) = \mathsf{a}(y)$ and, by symmetry, $W(y \mid -1) = \mathsf{a}(-y)$. Indeed, by using the assumption (5.2), we can show that this choice of $W(y \mid x)$ yields an L-density equal to $\mathsf{a}(y)$ (see Lemma 4.28 of [44]).

As a final reminder, the capacity $C(W)$ can be computed as a function of the L-density a according to the following formula (see Lemma 4.35 of [44]),

$$C(W) = \int \mathsf{a}(y) \left(1 - \log_2(1 + e^{-y})\right) dy. \tag{5.3}$$

5.2.2 How to Achieve the Symmetric Capacity of Asymmetric Channels

Problem Statement. Let W be a (not necessarily symmetric) B-DMC. The aim is to transmit over W with a rate close to $C_{\mathrm{s}}(W)$.

Design of the Scheme. The original construction of polar codes directly achieves the symmetric capacity of any B-DMC [37].

For sparse graph codes, some more analysis is required. Here, we will follow a line of reasoning inspired by Section 5.2 of [44]. A similar approach was first introduced in [172] and an alternative path that considers the average of the density evolution analysis with respect to each codeword, is considered in [173]. All these techniques lead to the same result.

The codebook of a code of block length N and rate R with parity check matrix $P \in \mathbb{F}_2^{(1-R)N \times N}$ is given by the set of $x_{1:N} \in \mathbb{F}_2^N$ such that $x_{1:N} P^T = 0_{1:(1-R)N}$, where $0_{1:(1-R)N}$ denotes a row vector of $(1-R)N$ zeros. In words, the transmitter and the receiver know that the results of the parity checks are all zeros. Let us consider a slightly different model in which the values of the parity checks are chosen uniformly at random and this randomness is shared between the transmitter and the receiver: first, we pick the parity checks uniformly at random; then, we pick a codeword uniformly at random among those that satisfy the parity checks. Clearly, this is equivalent to picking directly one codeword chosen uniformly at random from the whole space $\mathbb{F}_2^N$. As a result, we can model the codeword as a sequence of N uniform i.i.d. bits. Note that, in [172], instead of randomizing the cosets, the authors add a random scrambling vector to the entire codeword before transmission and then subtract it afterwards. The random scrambling and de-scrambling is absorbed into a normalized channel that is automatically symmetric, hence the standard density evolution equations hold. The concentration theorem (for a random code, scrambling vector, and channel realization) also follows by absorbing the scrambling bit into the randomness of the channel. This scrambling idea is also used in [174], where the authors explore the connection between symmetric channel coding, general channel coding, symmetric Slepian-Wolf coding, and general Slepian-Wolf coding.

As the channel can be asymmetric, we need to define two distinct L-densities according to the transmitted value. For simplicity, let us map the input alphabet $\mathbb{F}_2$

into $\{-1, 1\}$ and denote by $\mathsf{a}^+(y)$ and $\mathsf{a}^-(y)$ the L-density for the channel assuming that $X = 1$ and $X = -1$ is transmitted, respectively. Let us now flip the density associated with -1, i.e., we consider $\mathsf{a}^-(-y)$, so that positive values indicate "correct" messages. By the symmetry of the message-passing equations (see Definition 4.81 in [44]), the sign of all those messages that enter or exit the variable nodes with associated transmitted value -1 is flipped as well. Therefore, the density evolution analysis for a particular codeword is equivalent to that for the all-1 codeword, provided that we initialize the variable nodes with associated value 1 and -1 to $\mathsf{a}^+(y)$ and $\mathsf{a}^-(-y)$, respectively. Now, each transmitted bit is independent and uniformly distributed. Thus, we pick a variable node with L-density $\mathsf{a}^+(y)$ with probability $1/2$ and with L-density $\mathsf{a}^-(-y)$ with probability $1/2$. As a result, the density evolution equations for our asymmetric setting are the same as those for the transmission over the "symmetrized channel" with L-density given by

$$\mathsf{a}^{\mathrm{s}}(y) = \frac{1}{2}(\mathsf{a}^-(-y) + \mathsf{a}^+(y)). \tag{5.4}$$

This channel is, indeed, symmetric and its capacity equals the symmetric capacity of the actual channel W over which transmission takes place. These two results are formalized by the propositions below that are proved in Appendix 5.7.1.

Proposition 5.1. *Consider the transmission over a B-DMC and let* $\mathsf{a}^+(y)$ *and* $\mathsf{a}^-(y)$ *be the L-densities assuming that* $X = 1$ *and* $X = -1$ *is transmitted, respectively. Then, the L-density* $\mathsf{a}^{\mathrm{s}}(y)$ *given by* (5.4) *is symmetric.*

Proposition 5.2. *Consider the transmission over a B-DMC* W *with symmetric capacity* $C_{\mathrm{s}}(W)$ *and let* $\mathsf{a}^+(y)$ *and* $\mathsf{a}^-(y)$ *be the L-densities assuming that* $X = 1$ *and* $X = -1$ *is transmitted, respectively. Define the L-density of the "symmetrized channel"* $\mathsf{a}^{\mathrm{s}}(y)$ *as in* (5.4)*. Then,*

$$C_{\mathrm{s}}(W) = \int \mathsf{a}^{\mathrm{s}}(y) \left(1 - \log_2(1 + e^{-y})\right) dy. \tag{5.5}$$

Consequently, in order to achieve the symmetric capacity $C_{\mathrm{s}}(W)$ of the (possibly asymmetric) channel W, it suffices to construct a code that achieves the capacity of the symmetric channel with L-density a^{s}. Indeed, the density evolution analysis for the transmission over W is exactly the same as for the transmission over the symmetrized channel. Furthermore, the capacity of the symmetrized channel equals the symmetric capacity $C_{\mathrm{s}}(W)$ of the original channel W. As a result, in order to solve the problem, we can employ, for instance, an $(\mathtt{l}, \mathtt{r})$-regular SC-LDPC ensemble with sufficiently large degrees.

In short, the problem of achieving the symmetric capacity of any B-DMC can be solved by using codes (e.g., polar, spatially coupled) that are provably optimal for symmetric channels.

5.2.3 How to Transmit Biased Bits

Let us consider a generalization of the previous problem in which the bits of the codeword are biased, i.e., they are not chosen according to a uniform distribution. This scenario is an important primitive that will be used in Sections 5.4 and 5.5, where we describe coding techniques that achieve the capacity of asymmetric channels.

Problem Statement. Let W be a B-DMC with capacity-achieving input distribution $\{p^*(x)\}_{x\in\{0,1\}}$ such that $p^*(1) = \alpha$ for some $\alpha \in [0, 1]$. Let $X_{1:N}$ be a sequence of N i.i.d. Bernoulli(α) random variables. Denote by $Y_{1:N}$ the channel output when $X_{1:N}$ is transmitted. Furthermore, assume that the transmitter and the receiver are connected via a noiseless channel of capacity roughly $NH(X \mid Y)$. Given $Y_{1:N}$ and with the help of the noiseless channel, the aim is to reconstruct $X_{1:N}$ at the receiver with high probability as N goes large.

Design of the Scheme. Let P be a parity-check matrix with $NH(X \mid Y)$ rows and N columns. Hence, P represents a code of length N and rate $1-H(X \mid Y)$. Let $S_{1:NH(X|Y)} = X_{1:N}P^T$ and assume that $S_{1:NH(X|Y)}$ is sent over the noiseless channel to the receiver. Given $Y_{1:N}$ and $S_{1:NH(X|Y)}$, we will prove that the receiver can reconstruct $X_{1:N}$, assuming that the code represented by P is capacity-achieving under belief-propagation decoding for the symmetric channel described in the following (see (5.7)).

For the sake of simplicity, let us map the input alphabet into $\{-1, 1\}$. Since the input distribution is non-uniform, the belief-propagation (BP) algorithm needs to take into account also the prior on X and it is no longer based on the log-likelihood ratio $L(y)$ defined in (5.1). Let $L_{\mathrm{p}}(y)$ denote the log-posterior ratio, defined as

$$L_{\mathrm{p}}(y) = \ln \frac{p_{X|Y}(1 \mid y)}{p_{X|Y}(-1 \mid y)} = L(y) + \ln \frac{\bar{\alpha}}{\alpha}. \tag{5.6}$$

Following the lead of Section 5.2.2, let us define the densities of $L_{\mathrm{p}}(Y)$ assuming that $X = 1$ and $X = -1$ is transmitted and let us denote them as $\mathsf{a}_{\mathrm{p}}^{+}(y)$ and $\mathsf{a}_{\mathrm{p}}^{-}(y)$, respectively. If we flip the density associated with $X = -1$, i.e., we consider $\mathsf{a}_{\mathrm{p}}^{-}(-y)$, then, by the symmetry of the message-passing equations, the sign of the messages that enter or exit the variable nodes with associated transmitted value $X = -1$ is flipped as well. Therefore, the density evolution analysis for a particular codeword is equivalent to that for the all-one codeword provided that we initialize the variable nodes with associated value 1 to $\mathsf{a}_{\mathrm{p}}^{+}(y)$, and the variable nodes with value -1 to $\mathsf{a}_{\mathrm{p}}^{-}(-y)$, respectively. As $\mathbb{P}(X = -1) = \alpha$, the density evolution equations for our asymmetric setting are the same as those for the transmission over the "symmetrized channel" with L-density

$$\mathsf{a}_{\mathrm{p}}^{\mathrm{s}}(y) = \alpha \mathsf{a}_{\mathrm{p}}^{-}(-y) + \bar{\alpha} \mathsf{a}_{\mathrm{p}}^{+}(y). \tag{5.7}$$

The propositions below show that this channel is, indeed, symmetric and establish the relation between the conditional entropy $H(X \mid Y)$ and $\mathsf{a}_{\mathrm{p}}^{\mathrm{s}}(y)$. The proofs of these results can be found in Appendix 5.7.2.

Proposition 5.3. *The L-density $\mathsf{a}_{\mathrm{p}}^{\mathrm{s}}(y)$ given by (5.7) is symmetric.*

Proposition 5.4. *Consider the transmission over a B-DMC W with capacity-achieving input distribution p^*. Let $X \sim p^*$ and Y be the input and the output of the channel. Denote by $\mathsf{a}_{\mathrm{p}}^{+}(y)$ and $\mathsf{a}_{\mathrm{p}}^{-}(y)$ the densities of $L_{\mathrm{p}}(Y)$ assuming that $X = 1$ and $X = -1$ is transmitted. Let $\mathsf{a}_{\mathrm{p}}^{\mathrm{s}}(y)$ be the density of the "symmetrized channel", as in (5.7). Then,*

$$H(X \mid Y) = \int \mathsf{a}_{\mathrm{p}}^{\mathrm{s}}(y) \log_2(1 + e^{-y}) dy. \tag{5.8}$$

Let us now see how these two propositions imply that we can reconstruct $X_{1:N}$ with high probability. Since the channel with density $\mathsf{a}_{\mathrm{p}}^{\mathrm{s}}(y)$ is symmetric by Proposition 5.3, its capacity is given by

$$\int \mathsf{a}_{\mathrm{p}}^{\mathrm{s}}(y)\left(1-\log_2(1+e^{-y})\right)dy = 1-\int \mathsf{a}_{\mathrm{p}}^{\mathrm{s}}(y)\log_2(1+e^{-y})dy = 1-H(X\mid Y),$$

where the last equality comes from Proposition 5.4. Recall that the receiver is given the channel output $Y_{1:N}$ and the error-free vector $S_{1:NH(X|Y)} = X_{1:N}P^T$. Hence, we can think of this setting as one in which $X_{1:N}$ is a codeword of a sparse graph code with syndrome vector $S_{1:NH(X|Y)}$ shared between the transmitted and the receiver. As previously stated, the density evolution analysis for this case is the same as when we transmit over the symmetric channel with density $\mathsf{a}_{\mathrm{p}}^{\mathrm{s}}(y)$ and the syndrome vector is set to 0. By assumption, the matrix P comes from a code that achieves capacity for such a symmetric channel, hence the transmitted vector $X_{1:N}$ can be reconstructed with high probability.

We can employ, for instance, an $(\mathtt{l},\mathtt{r})$-regular SC-LDPC ensemble with sufficiently large degrees. Another option is to use spatially coupled MacKay-Neal (MN) and Hsu-Anastasopoulos (HA) LDPC codes that, compared to $(\mathtt{l},\mathtt{r})$-regular codes, have bounded graph density. In particular, in [175] it is proved that MN and HA codes achieve the capacity of B-DMCs *under MAP decoding* by using a parity-check matrix with bounded column and row weight. Furthermore, the authors of [175] give empirical evidence of the fact that spatially coupled MN and HA codes achieve the capacity of the BEC also *under iterative decoding*. These results are extended to the additive white Gaussian noise channel in [176].

In summary, so far we have discussed how to achieve the symmetric capacity of a B-DMC and how to transmit biased bits. Now, let us move to the main topic of this paper and describe three approaches for achieving the actual capacity of any DMC. While doing so, we will regard the solutions to the two problems of this section as useful primitives.

5.3 Paradigm 1: Gallager's Mapping

The solution proposed by Gallager in page 208 of [120] consists of using a standard linear code and applying a non-linear mapper to the encoded bits in such a way that the resulting input distribution is appropriately biased. More recently, Gallager's mapping was used in [40] to approach the capacity of nonstandard channels via turbo-like codes. Furthermore, in [177], the authors applied this shaping idea to finite-state channels and described how to construct an explicit invertible finite-state encoder. Before moving on to a general description of the scheme, to convey the main ideas, let us start with an example.

5.3.1 A Concrete Example

Let $\mathcal{X}=\{0,1,2\}$ and suppose that we want to transmit over a channel $W:\mathcal{X}\to\mathcal{Y}$ with a capacity-achieving input distribution of the following form: $p^*(0)=3/8$, $p^*(1)=3/8$, $p^*(2)=2/8$.

Let $\mathcal{V}=\{0,1,\cdots,7\}$ and consider the function $f:\mathcal{V}\to\mathcal{X}$ that maps three elements of $\mathcal{V}$ into $0\in\mathcal{X}$, three other elements of $\mathcal{V}$ into $1\in\mathcal{X}$, and the remaining

two elements of $\mathcal{V}$ into $2 \in \mathcal{X}$. In this way, the uniform distribution over $\mathcal{V}$ induces the capacity-achieving distribution over $\mathcal{X}$. Define the channel $W' : \mathcal{V} \to \mathcal{Y}$ as

$$W'(y \mid v) = W(y \mid f(v)). \tag{5.9}$$

Take a code that achieves the symmetric capacity of W'. Then, we can use this code to achieve the capacity of W via the mapping f.

The above scheme works under the assumption that we can construct codes that achieve the symmetric capacity for any given input alphabet size. Note that this can be done, e.g., with q-ary polar codes [79]. Sometimes it is more convenient to achieve this goal indirectly by using only binary codes. Indeed, suppose that the channel changes for some reason. Then, the optimal input distribution also changes, and we might have to change the alphabet $\mathcal{V}$. If we code directly on $\mathcal{V}$, we will also have to change the code itself. If the code needs to be implemented in hardware, this might not be convenient. However, if we manage to use the same binary code and only modify some preprocessing steps, then it is easy to accomplish any required change in the input distribution.

Let us now describe this approach in detail. Observe that $\mathcal{V}$ has cardinality $8 = 2^3$. Rather than considering the set of integers from 0 to 7, it is more convenient to consider the set of binary triplets. Let $\mathcal{U} = \{0,1\}^3$ and consider the function by $g : \mathcal{U} \to \mathcal{X}$. As before, g maps three elements of $\mathcal{U}$ into $0 \in \mathcal{X}$, three other elements of $\mathcal{U}$ into $1 \in \mathcal{X}$, and the remaining two elements of $\mathcal{U}$ into $2 \in \mathcal{X}$. In this way, the uniform distribution over $\mathcal{U}$ induces the capacity-achieving distribution over $\mathcal{X}$. Note that any $u \in \mathcal{U}$ can be written as $u = (u^{(1)}, u^{(2)}, u^{(3)})$, where $u^{(i)} \in \{0,1\}$ for $i \in \{1,2,3\}$. Define the channels $W_1'' : \{0,1\} \to \mathcal{Y}$, $W_2'' : \{0,1\} \to \mathcal{Y} \times \{0,1\}$, $W_3'' : \{0,1\} \to \mathcal{Y} \times \{0,1\} \times \{0,1\}$ as

$$\begin{aligned}
W_1''(y \mid u^{(1)}) &= \frac{1}{4} \sum_{u^{(2)}, u^{(3)}} W(y \mid g(u^{(1)}, u^{(2)}, u^{(3)})), \\
W_2''(y, u^{(1)} \mid u^{(2)}) &= \frac{1}{4} \sum_{u^{(3)}} W(y \mid g(u^{(1)}, u^{(2)}, u^{(3)})), \\
W_3''(y, u^{(1)}, u^{(2)} \mid u^{(3)}) &= \frac{1}{4} W(y \mid g(u^{(1)}, u^{(2)}, u^{(3)})).
\end{aligned} \tag{5.10}$$

Take three binary codes that achieve the symmetric capacities of W_1'', W_2'', and W_3''. By the chain rule of mutual information, the sum of these capacities equals $C(W)$. Hence, we can use these codes to achieve the capacity of W via the mapping g.

5.3.2 Description of the General Scheme

Problem Statement. Let W be a DMC with capacity-achieving input distribution $\{p^*(x)\}_{x \in \mathcal{X}}$. The aim is to transmit over W with rate close to $C(W)$.

Design of the Scheme. Pick $\delta > 0$ and find a rational approximation $\tilde{p}(x)$ that differs from $p^*(x)$ by at most δ in total variation distance. In formulae, take $\tilde{p}(x) = N_x/d$ with N_x, $d \in \mathbb{N}$ for all $x \in \mathcal{X}$ such that

$$\frac{1}{2} \sum_{x \in \mathcal{X}} |p^*(x) - \tilde{p}(x)| < \delta. \tag{5.11}$$

Take an extended alphabet $\mathcal{V}$ with cardinality equal to d and consider the function $f : \mathcal{V} \to \mathcal{X}$ that maps N_x elements of $\mathcal{V}$ into $x \in \mathcal{X}$. Define the channel $W' : \mathcal{V} \to \mathcal{Y}$ as in (5.9). Denote by X and Y the input and the output of the channel W, respectively. Let V be uniformly distributed over $\mathcal{V}$ and set $X = f(V)$. Since the uniform distribution over $\mathcal{V}$ induces the input distribution $\tilde{p}(x)$ over $\mathcal{X}$, we have that $X \sim \tilde{p}(x)$. Construct a code $\mathcal{C}$ that achieves the symmetric capacity of W'. Therefore, by using the code $\mathcal{C}$ and the mapping f, we can transmit at rate R arbitrarily close to

$$C_{\mathrm{s}}(W') = I(V;Y) = I(X;Y) \underset{\delta \to 0}{\to} C(W).$$

As δ goes to 0, the distribution $\tilde{p}$ tends to p^* and $I(X;Y)$ approaches $C(W)$.

If we want to restrict to binary codes, select a rational approximation of the form $\tilde{p}(x) = N_x/2^t$ for $t, N_x \in \mathbb{N}$. Pick $\mathcal{U} = \{0,1\}^t$ and consider the function $g : \mathcal{U} \to \mathcal{X}$ that maps N_x elements of $\mathcal{U}$ into $x \in \mathcal{X}$. The set $\mathcal{U}$ contains binary vectors of length t that can be written in the form $u^{1:t} = (u^{(1)}, \cdots, u^{(t)})$, where $u^{(j)} \in \{0,1\}$ for $j \in [t]$. Define the synthetic channels $W''_j : \{0,1\} \to \mathcal{Y} \times \{0,1\}^{j-1}$, similarly to (5.10), i.e.,

$$W''_j(y, u^{1:j-1} \mid u^{(j)}) = \frac{1}{2^{t-1}} \sum_{u^{j+1:t}} W(y \mid g(u^{1:t})). \tag{5.12}$$

Let $U_{1:t}$ be a sequence of t i.i.d. random variables uniform over $\{0,1\}$. Set $X = g(U_{1:t})$. Since the uniform distribution over $\mathcal{U}$ induces the input distribution $\tilde{p}(x)$ on $\mathcal{X}$, we have that $X \sim \tilde{p}(x)$. Construct t codes $\mathcal{C}_1, \cdots, \mathcal{C}_t$ such that $\mathcal{C}_j$ has rate R_j that is arbitrarily close to the symmetric capacity of the channel W''_j. Therefore, by using these codes and the mapping g, we can transmit at rate R arbitrarily close to

$$\sum_{j=1}^{t} I(U^{(j)}; Y \mid U_{1:j-1}) = I(X;Y) \underset{t \to \infty}{\to} C(W), \tag{5.13}$$

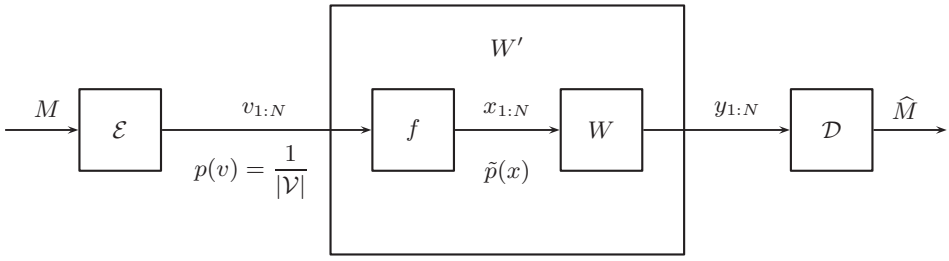
where the first equality comes from the chain rule and $I(X;Y)$ approaches $C(W)$ as δ goes to 0.

Let us now explain formally how the encoding and decoding operations are done for the schemes mentioned above (see also Figure 5.2). Then, we will consider the performance of this approach by relating the gap $C(W) - I(X;Y)$ to δ and to the cardinalities of the input and the output alphabets.

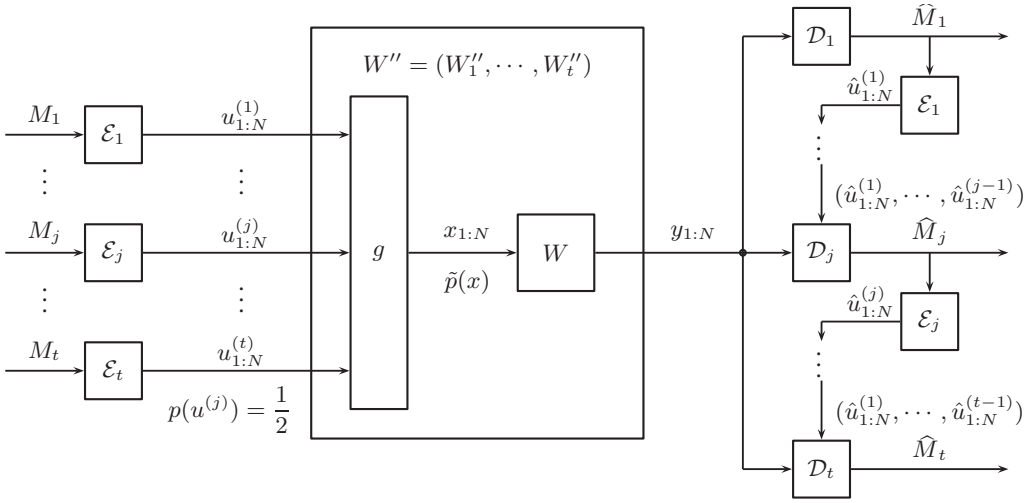
Encoding. First, consider the scheme based on a single non-binary code. Let M be the information message that can be thought of as a binary string of length nR and let $\mathcal{E}$ be the encoder of the code $\mathcal{C}$. The output of the encoder is $v_{1:N} = (v_1, \cdots, v_N)$, where $v_i \in \mathcal{V}$ for $i \in [N]$. Then, $v_{1:N}$ is mapped component-wise by the function f into $x_{1:N} = (x_1, \cdots, x_N)$, with $x_i \in \mathcal{X}$ such that $x_i = f(v_i)$.

Second, consider the scheme based on t binary codes. Let $M = (M_1, \cdots, M_t)$ be the information message divided into t parts so that M_j can be thought of as a binary string of length nR_j for $j \in \{1, \cdots, t\}$. Let $\mathcal{E}_j$ be the encoder of the code $\mathcal{C}_j$ that maps M_j into $u^{(j)}_{1:N} = (u^{(j)}_1, \cdots, u^{(j)}_N)^T$, where $u^{(j)}_i \in \{0,1\}$ for $i \in [N]$. Then, $u^{1:t}_{1:N}$ is mapped component-wise by the function g into $x_{1:N} = (x_1, \cdots, x_N)$, with $x_i \in \mathcal{X}$ given by $x_i = g(u^{1:t}_i)$.

Finally, we transmit the sequence $x_{1:N}$ over the channel W.



(a) Solution based on a single non-binary code: the message M is encoded by $\mathcal{E}$ and decoded by $\mathcal{D}$.



(b) Solution based on t binary codes: the message $M = (M_1, \cdots, M_t)$ is encoded by $\mathcal{E}_1, \cdots, \mathcal{E}_t$ and decoded successively by $\mathcal{D}_1, \cdots, \mathcal{D}_t$. Note that $\mathcal{D}_j$ is the decoder of the synthetic channel W_j'' and it is fed with the output of the actual channel W together with the previous re-encoded estimates.

Figure 5.2 – Coding over asymmetric channels via Gallager's mapping.

Decoding. First, consider the scheme based on a single non-binary code. Let $\mathcal{D}$ be the decoder of the code $\mathcal{C}$, that accepts as input the channel output $y_{1:N}$ and outputs the estimate $\widehat{M}$.

Second, consider the scheme based on t binary codes. Let $\mathcal{D}_j$ be the decoder of the code $\mathcal{C}_j$. It accepts as input the channel output $y_{1:N}$ and the previous re-encoded estimates $(\widehat{u}_{1:N}^{(1)}, \cdots, \widehat{u}_{1:N}^{(j-1)})$. It outputs the current estimate $\widehat{M}_j$. To make the use of the previous estimates possible, the decoding occurs successively, i.e., the decoders $\mathcal{D}_1, \cdots, \mathcal{D}_t$ are activated in series.

The situation is schematically represented in Figure 5.2.

Performance. The codes $\mathcal{C}$ and $\mathcal{C}_j$ can be used to transmit reliably at rates R and R_j. Then, $\widehat{M} = M$ and $\widehat{M}_j = M_j$ $(j \in [t])$ with high probability. As a result, we can transmit over W with rate close to $I(X;Y)$, where the input distribution is $\tilde{p}(x)$. Also, as the mutual information is a continuous function of the input distribution,

if δ gets small, then $I(X;Y)$ approaches $C(W)$. This statement is made precise by the following proposition that is proved in Appendix 5.7.3.

Proposition 5.5. *Consider the transmission over the channel $W : \mathcal{X} \to \mathcal{Y}$ and let $I(p)$ be the mutual information between the input and the output of the channel when the input distribution is p. Let p and p^* be input distributions such that their total variation distance is upper bounded by δ, as in* (5.11)*, for $\delta \in (0, 1/8)$. Then,*

$$|I(p^*) - I(p)| < 3\delta \log_2 |\mathcal{Y}| + h_2(\delta), \tag{5.14}$$

$$|I(p^*) - I(p)| < 7\delta \log_2 |\mathcal{X}| + h_2(\delta) + h_2(4\delta). \tag{5.15}$$

Note that the bounds (5.14) and (5.15) depend separately on the input and the output alphabet. Therefore, we can conclude that, under the hypotheses of Proposition 5.5,

$$|I(p^*) - I(p)| = O\left(\delta \log_2 \left(\frac{\min(|\mathcal{X}|, |\mathcal{Y}|)}{\delta}\right)\right).$$

5.4 Paradigm 2: Integrated Scheme

The basic idea of this approach is to use a coding scheme that is simultaneously good for lossless source coding and for channel coding. The *source coding* part is needed to create a biased input distribution from uniform bits, whereas the *channel coding* part provides reliability for the transmission over the channel.

A provably capacity-achieving scheme was first proposed in [121] in the context of polar codes. We reviewed such a scheme in Section 4.4.2, because we used it as a primitive to design polar codes for the broadcast channel. Let us now briefly clarify how the scheme of [121] solves both a source coding and a channel coding problem.

As for *source coding*, we consider the sets $\mathcal{H}_X$ and $\mathcal{L}_X$ defined in (4.10): for $i \in \mathcal{H}_X$, the bit U_i is approximately uniformly distributed and independent of $U_{1:i-1}$; and, for $i \in \mathcal{L}_X$, the bit U_i is approximately a deterministic function of $U_{1:i-1}$. Consequently, we can compress $X_{1:N}$ into the sequence $\{U_i\}_{i \in \mathcal{L}_X^c}$ that has size roughly $NH(X)$ (for a more detailed explanation, see also Section 4.4.1).

As for *channel coding*, we interpret the channel output $Y_{1:N}$ as side information for $X_{1:N}$ and consider the sets $\mathcal{H}_{X|Y}$ and $\mathcal{L}_{X|Y}$ defined in (4.15) and (4.16): for $i \in \mathcal{H}_{X|Y}$, U_i is approximately uniformly distributed and independent of $(U_{1:i-1}, Y_{1:N})$; and, for $i \in \mathcal{L}_{X|Y}$, U_i is approximately a deterministic function of $(U_{1:i-1}, Y_{1:N})$.

To construct a polar code, we place the information in the positions indexed by $\mathcal{I} = \mathcal{H}_X \cap \mathcal{L}_{X|Y}$. Indeed, if $i \in \mathcal{I}$, U_i is approximately uniformly distributed given $U_{1:i-1}$, which implies that U_i is suitable to contain information. Furthermore, U_i is approximately a deterministic function given $U_{1:i-1}$ *and* $Y_{1:N}$, which implies that U_i is decodable. From (4.19), we obtain that $\mathcal{I}$ has size roughly $NI(X;Y)$. Hence, by choosing X to be distributed according to the capacity-achieving input distribution p^*, we can transmit over W with rate close to $C(W)$.

In the remaining part of this section, we describe how to extend this approach to sparse graph codes.

Problem Statement. Let W be a B-DMC with capacity-achieving input distribution $\{p^*(x)\}_{x\in\{0,1\}}$ such that $p^*(1) = \alpha$ for some $\alpha \in [0, 1]$. The aim is to transmit over W with rate close to $C(W)$.

Design of the Scheme. Consider a linear code with parity-check matrix P with $NH(X) = Nh_2(\alpha)$ rows and N columns, namely $P \in \mathbb{F}_2^{Nh_2(\alpha)\times N}$. Let $X_{1:N} \in \mathbb{F}_2^N$ be a codeword and denote by $Y_{1:N}$ the corresponding channel output. Let $S_{1:Nh_2(\alpha)} \in \mathbb{F}_2^{Nh_2(\alpha)}$ be the vector of syndromes defined as $S_{1:Nh_2(\alpha)} = X_{1:N}P^T$.

Recall that, in the integrated scheme, we need to achieve the source coding and the channel coding part at the same time. To do so, we divide $S_{1:Nh_2(\alpha)}$ into two parts, i.e.,

$$S_{1:Nh_2(\alpha)} = (S^{(1)}_{1:NC(W)}, S^{(2)}_{1:NH(X|Y)}), \tag{5.16}$$

where this decomposition is possible because $h_2(\alpha) = H(X) = C(W) + H(X \mid Y)$. Similarly, it is convenient to write the parity-check matrix P as

$$P^T = [P_1^T, P_2^T], \tag{5.17}$$

where $P_1 \in \mathbb{F}_2^{NC(W)\times N}$ and $P_2 \in \mathbb{F}_2^{NH(X|Y)\times N}$.

The first part of the decomposition (5.16), namely $S^{(1)}_{1:NC(W)}$, contains the information bits. This is quite different from what happens in a standard parity-check code, in which the values of the parity checks are shared between the encoder and the decoder (and typically fixed to 0). In the proposed scheme, the parity checks contain the transmitted message.

The second part, namely $S^{(2)}_{1:NH(X|Y)}$, is chosen uniformly at random, and this randomness is assumed to be shared between the transmitter and the receiver. Note that $S^{(2)}_{1:NH(X|Y)}$ does not depend on the information bits.

The choice of the parity-check matrix P_2 concerns the *channel coding* part of the scheme. Recall the problem considered in Section 5.2.3: given the channel output $Y_{1:N}$ and the parity bits $S^{(2)}_{1:NH(X|Y)}$, the receiver can reconstruct $X_{1:N}$, as long as the parity-check matrix corresponds to a code that achieves the capacity of the "symmetrized channel" with density given by (5.7). For example, we can set P_2 to be the parity-check matrix of an $(\mathtt{l}, \mathtt{r})$-regular SC-LDPC ensemble with sufficiently large degrees.

The choice of the parity-check matrix P_1 concerns the *source coding* part of the scheme. In particular, we choose P_1 in order to fulfill the following requirement: we want to associate with each syndrome $S_{1:Nh_2(\alpha)}$ a codeword $X_{1:N}$ with $X_{1:N}[P_1^T, P_2^T] = S_{1:Nh_2(\alpha)}$ so that the uniform i.i.d. distribution on the syndromes induces a Bernoulli(α) i.i.d. distribution on the codewords.

Before moving on with the description of the scheme, let us review how to use sparse graph codes to accomplish lossless source coding. We are given a vector $X_{1:N}$ of N i.i.d. Bernoulli(α) random variables and the aim is to compress it into a binary sequence of size roughly $Nh_2(\alpha)$. We want to solve the problem by using the parity-check matrix $\tilde{P} \in \mathbb{F}_2^{Nh_2(\alpha)\times N}$ of a sparse graph code as the linear compressor and the BP decoder as the decompressor, respectively [171]. More specifically, given $x_{1:N}$ to be compressed, the encoder computes $s_{1:Nh_2(\alpha)} = x_{1:N}\tilde{P}^T$. The task of the decoder can be summarized as follows:

Task 1. Given the syndrome vector $x_{1:N}\tilde{P}^T$, recover the biased vector $x_{1:N}$ by using the BP algorithm.

Les us now relate this task to a channel coding problem. Let $c_{1:N}$ be a codeword of the code with parity-check matrix $\tilde{P}$, i.e., $c_{1:N}\tilde{P}^T = 0_{1:Nh_2(\alpha)}$. Consider the transmission of $c_{1:N}$ over the binary symmetric channel with crossover probability α, i.e., the BSC(α), and let $y_{1:N}$ be the channel output. Denote by $y_{1:N}\tilde{P}^T$ the syndrome computed by the decoder, and note that $y_{1:N}\tilde{P}^T = e_{1:N}\tilde{P}^T$, where $e_i = 1$ if the i-th bit was flipped by the channel, and 0 otherwise. Consider the following two tasks:

Task 2. Given the syndrome vector $e_{1:N}\tilde{P}^T$, recover the error vector $e_{1:N}$ by using the BP algorithm.

Task 3. Given the received vector $y_{1:N}$, recover the transmitted codeword $c_{1:N}$ by using the BP algorithm.

Let us briefly show that these three tasks are, in fact, equivalent. First of all, note that *Task 1* and *Task 2* are clearly identical. Furthermore, it is shown in [171] that *Task 2* succeeds if and only if *Task 3* succeeds. The idea is to write down the message-passing equations in the two cases, and to observe that the messages obtained in *Task 2* can be put in one-to-one correspondence with the messages obtained in *Task 3*. More specifically, on the one hand, in *Task 2* we initialize all the received values at variable nodes by $\ln(\bar{\alpha}/\alpha)$ and the check nodes have an associated sign given by the vector $(-1)^{e_{1:N}\tilde{P}^T}$. On the other hand, in *Task 3*, we initialize the received values at variable nodes by the vector $(-1)^{y_{1:N}} \cdot \ln(\bar{\alpha}/\alpha)$, and all the check nodes have an associated sign of $+1$. The crucial observation is that, for each iteration of the BP algorithm, the modulus of the received values at variable nodes stays the same for the two tasks, and the sign is flipped according to the value of $y_{1:N}$.

Note that *Task 3* is the standard channel coding problem for the transmission over the BSC. Hence, we can use the parity-check matrix of a code that achieves capacity over the BSC to compress N i.i.d. Bernoulli(α) random variables into a binary sequence of size roughly $Nh_2(\alpha)$.

Let us come back to our original problem of achieving the capacity of a B-DMC. The source coding part of our approach is basically the inverse of source coding. Indeed, given the uniform vector of syndromes $S_{1:Nh_2(\alpha)}$, we want to obtain a biased codeword $X_{1:N}$.

Let P_1 be the parity-check matrix of a regular SC-LDPC ensemble with sufficiently large degrees. This implies that also P is the parity-check matrix of a regular SC-LDPC ensemble. First, suppose that the vector of syndromes to be fulfilled has size m slightly larger than $Nh_2(\alpha)$, say $m = N(h_2(\alpha) + \epsilon)$ for some small $\epsilon > 0$. Consequently, suppose that the matrix P has m rows. From the argument above, if $X_{1:N}$ is a vector of N i.i.d. Bernoulli(α) random variables, then, with high probability, there exists a vector of syndromes $S_{1:m}$ such that $X_{1:N}P^T = S_{1:m}$. However, only for a vanishing fraction of possible $S_{1:m}$ there exists $X_{1:N}$ such that $X_{1:N}P^T = S_{1:m}$. This means that, for a randomly chosen $S_{1:m}$, with high probability the BP algorithm will not succeed.

Suppose now that the vector of syndromes to be fulfilled has size m no larger than $Nh_2(\alpha)$. Then, with high probability, there are exponentially many $X_{1:N}$ with i.i.d. Bernoulli(α) distribution such that $X_{1:N}P^T = S_{1:m}$. This implies that the

BP algorithm does not converge, as a message-passing decoder operating locally can easily get confused when there are many feasible solutions.

Perhaps a more apt approach is to frame the source coding part of our scheme as a lossy compression problem, where the distortion between the distribution of $X_{1:N}$ and an i.i.d. Bernoulli(α) distribution tends to 0 as N goes large. It was observed in [178] that using a standard BP algorithm is not effective for lossy compression, and that this issue can be overcome by introducing a decimation process. An encoding scheme for lossy compression based on spatially coupled low-density generator-matrix (LDGM) codes and belief-propagation guided decimation is presented in [179], where it is shown with numerical simulations that the spatially coupled ensemble approaches the Shannon rate-distortion limit for large check degrees. This technique is extended to the Wyner-Ziv and Gelfand-Pinsker problems in [180], where it is shown empirically that spatially coupled compound LDGM/LDPC codes with belief-propagation guided decimation achieve the optimal rates. In particular, the solution to the Gelfand-Pinsker problem presents some similarities to our approach: the information bits are placed in a vector of syndromes, and the compound LDGM/LDPC codes are simultaneously good for rate distortion and channel coding. The need for a scheme that is good both for source and channel coding is due to the fact that, in the Gelfand-Pinsker setting, there is a constraint on the average weight of the transmitted codeword. This is analogous to our requirement that $X_{1:N}$ has a Bernoulli(α) i.i.d. distribution. Another solution to the Gelfand-Pinsker problem that adopts a framework similar to the one considered in this section is provided in [181]. Here, the authors use LDPC matrices with logarithmic column weight and maximum likelihood decoding. This approach provably achieves the optimal rate by introducing the notion of a hash property. However, the decoding algorithm has exponential complexity. The results of [181] are extended in [182], where codes for general (thus, possibly asymmetric) channels and sources are constructed. It is interesting to point out that the problem of generating the codeword $X_{1:N}$ is solved in [182] with a constrained-random-number generator, instead of resorting to a belief-propagation type of algorithm.

Our solution follows the lead of [179, 180] and uses belief-propagation guided decimation at the encoder. Note that this approach works well in practice, as testified by the simulation results in [179,180], but we currently have no theoretical guarantees on its performance. Let us now get down to the details of the proposed encoding scheme. Given the syndrome vector $S_{1:Nh_2(\alpha)}$, we run the standard BP algorithm and, after every t iterations, for some fixed $t \in \mathbb{N}$, we decimate a small fraction of the codeword bits. This means that we set each decimated bit to its most likely value. Furthermore, we fix the modulus of the received values at the corresponding variable nodes to $+\infty$. Consequently, the decimated bits will not change during the next iterations of the algorithm. The procedure ends when all the codeword bits have been decimated.

The algorithm described above outputs the codeword $X_{1:N}$. Recall that $S^{(1)}_{1:NC(W)}$ contains the information bits. Hence, if $X_{1:N}P_1^T$ differs from $S^{(1)}_{1:NC(W)}$, even if the decoder correctly reconstructs the transmitted codeword $X_{1:N}$, it will not correctly reconstruct the information sequence. However, the idea is that the fraction of positions in which $X_{1:N}P_1^T$ differs from $S^{(1)}_{1:NC(W)}$ tends to 0 as N goes large. This intuition is confirmed by the numerical simulations in [179, 180]. Hence, in order to

cope with this issue, we pre-code $S^{(1)}_{1:NC(W)}$ with a negligible loss in rate.

Encoding. First, we pre-code the $NC(W)$ information bits with a code $\mathcal{C}_{\mathrm{p}}$ of rate close to 1. We can use, for instance, an SC-LDPC code or a polar code designed for the transmission over the BSC. The output of this pre-coding operation is the sequence $s^{(1)}_{1:N(C(W)+\epsilon)}$, for some small $\epsilon > 0$. Then, we fill $s^{(2)}_{1:NH(X|Y)}$ with a realization of a sequence chosen uniformly at random and shared between the transmitter and the receiver. Let P be the parity-check matrix of an SC-LDPC code with sufficiently large degrees. From the syndrome vector $s_{1:N(H(X)+\epsilon)} = (s^{(1)}_{1:N(C(W)+\epsilon)}, s^{(2)}_{1:NH(X|Y)})$ and the parity-check matrix P, we obtain the codeword $x_{1:N}$ by running the BP algorithm with decimation steps.

Let P_1 be obtained by decomposing P as in (5.16). Let us now check that we can recover correctly the initial information bits from $x_{1:N}P_1^T$ by decoding $\mathcal{C}_{\mathrm{p}}$. If this is not possible, then the overall procedure is repeated with a different choice for the code $\mathcal{C}_{\mathrm{p}}$. Once the decoding of $\mathcal{C}_{\mathrm{p}}$ succeeds, we transmit the vector $x_{1:N}$ over the channel.

Decoding. The decoder receives $y_{1:N}$ and runs the BP algorithm using also the vector of syndromes $s^{(2)}_{1:NH(X|Y)}$ shared with encoder. Let $\hat{x}_{1:N}$ be the output of the BP algorithm. Eventually, we recover the information bits from $\hat{x}_{1:N}P_1^T$ by decoding $\mathcal{C}_{\mathrm{p}}$.

Performance. There are two possible types of errors. First, the encoder might fail to produce a codeword $x_{1:N}$ for which the decoding of $\mathcal{C}_{\mathrm{p}}$ succeeds. Second, the decoder might not estimate correctly the transmitted vector, i.e., $\hat{x}_{1:N} \neq x_{1:N}$. Note that, by construction, if $\hat{x}_{1:N} = x_{1:N}$, then the decoder recovers correctly the information bits.

The second error event occurs with vanishing probability, and this result is provable by following the argument of Section 5.2.3. Concerning the first error event, we only need that it does not happen with probability 1, because the encoding operation can be attempted several times. As we previously pointed out, from numerical simulations we observe that $x_{1:N}P_1^T$ agrees with $s^{(1)}_{1:N(C(W)+\epsilon)}$ in almost all the positions. We remark that, to the best of our knowledge, this last statement is not provable, because of the decimation steps introduced in the BP algorithm. If $x_{1:N}P_1^T$ and $s^{(1)}_{1:N(C(W)+\epsilon)}$ are sufficiently close, then, with high probability, we can recover the information bits by decoding $\mathcal{C}_{\mathrm{p}}$.

5.5 Paradigm 3: Chaining Construction

In the integrated approach, discussed in the preceding section, the idea was to use a single code to accomplish both the source and the channel coding part. The chaining construction, on the contrary, enables us to separate these two tasks. In this way, we can combine *any* solution to the source coding part with *any* solution to the channel coding part. This idea was first presented in [122]. Here, we prove that it can be used to achieve the capacity of any asymmetric DMC.

The problem statement is the same as in Section 5.4. Our main idea is formalized by the following theorem.

Theorem 5.1 (Chaining Construction for Asymmetric Channels)**.** *Let W be a B-DMC with capacity-achieving input distribution $\{p^*(x)\}_{x\in\{0,1\}}$ such that $p^*(1) = \alpha$ for some $\alpha \in [0,1]$. Denote by X and Y the input and the output of the channel, respectively. Let $N, m, \ell \in \mathbb{N}$, where m is roughly $Nh_2(\alpha)$, and ℓ is roughly $NH(X \mid Y)$. Denote by $U_{1:m}$ a sequence of m i.i.d. uniform random variables, and by $X_{1:N}$ a sequence of N i.i.d. Bernoulli(α) random variables. Let $Y_{1:N}$ be the channel output when $X_{1:N}$ is transmitted. Assume that, for any $\delta > 0$, there exists $N \in \mathbb{N}$ and there exist maps $f : \{0,1\}^N \to \{0,1\}^m$, $g : \{0,1\}^m \to \{0,1\}^N$, and $h : \{0,1\}^N \to \{0,1\}^\ell$ that satisfy the following properties.*

1. *$U_{1:m} = f(g(U_{1:m}))$, i.e., the map f is invertible, with probability $1 - \delta$.*
2. *The total variation distance between the distribution of $g(U_{1:m})$ and the distribution of $X_{1:N}$ is upper bounded by δ.*
3. *Given $Y_{1:N}$ and $h(X_{1:N})$, it is possible to reconstruct $X_{1:N}$ with probability $1 - \delta$.*

Then, we can use f, g, and h to transmit over W with rate close to $C(W)$.

In the following, we will prove this theorem and we will provide choices of f, g, and h that fulfill the required properties.

Design of the Scheme. First, we consider the *source coding* part of the scheme. Recall that in the previous section we framed this task as the inverse of source coding and we described a solution that uses sparse graph codes and belief-propagation guided decimation. Let us now consider this problem from a more general point of view.

In the traditional lossless compression setting, the input is a sequence $X_{1:N}$ with i.i.d. Bernoulli(α) distribution, and the encoder consists of a map from the set $\{0,1\}^N$ of source sequences to the set $\{0,1\}^*$ of finite-length binary strings. Let $f : \{0,1\}^N \to \{0,1\}^*$ be the encoding map, so that $U = f(X_{1:N})$ is the compressed description of the source sequence $X_{1:N}$. For a good source code, the expected binary length of U is close to the entropy of the source, i.e., $Nh_2(\alpha)$. In addition, the decoding function $g : \{0,1\}^* \to \{0,1\}^N$ is such that $X_{1:N} = g(f(X_{1:N}))$ with high probability over the choice of $X_{1:N}$. Several solutions to this problem have been proposed to date, such as, Huffman coding, arithmetic coding, Lempel-Ziv compression [183], polar codes [91, 92], and LDPC codes [171], just to name a few.

In our setting, the input is the compressed sequence $U_{1:m}$ with i.i.d. uniform distribution, instead of the biased sequence $X_{1:N}$. Note that $U_{1:m}$ contains the information bits. Furthermore, we consider maps from $\{0,1\}^N$ to $\{0,1\}^m$ and vice versa, where m is a fixed integer of size roughly $Nh_2(\alpha)$, instead of maps from $\{0,1\}^N$ to $\{0,1\}^*$. More specifically, the encoder and the decoder implement the maps $g : \{0,1\}^m \to \{0,1\}^N$ and $f : \{0,1\}^N \to \{0,1\}^m$, respectively. For the source coding part of our scheme, we require that the maps f and g satisfy the first two properties stated in Theorem 5.1. Let us justify such requirements.

The first property ensures that, given $g(U_{1:m})$, it is possible to recover $U_{1:m}$ with high probability. This requirement is crucial because $g(U_{1:m})$ represents the codeword transmitted over the channel. Hence, at the decoder, given the channel output, we estimate $g(U_{1:m})$ and, from this, we deduce the information vector $U_{1:m}$.

The second property ensures that the error probability for the transmission of $g(U_{1:m})$ is upper bounded by the error probability for the transmission of N i.i.d. Bernoulli(α) random variables plus δ. This statement can be proved as follows. Recall that, by definition of total variation distance, the second property can be written as

$$\frac{1}{2}\sum_{x\in\{0,1\}^N} |\mathbb{P}_{g(U_{1:m})}(x) - \mathbb{P}_{X_{1:N}}(x)| < \delta, \tag{5.18}$$

where $X_{1:N}$ has an i.i.d. Bernoulli(α) distribution. Then, by using that

$$\sum_{x\in\{0,1\}^N} \mathbb{P}_{g(U_{1:m})}(x) = \sum_{x\in\{0,1\}^N} \mathbb{P}_{X_{1:N}}(x) = 1,$$

we obtain

$$\sum_{x\in\{0,1\}^N} \max(\mathbb{P}_{g(U_{1:m})}(x) - \mathbb{P}_{X_{1:N}}(x), 0) < \delta. \tag{5.19}$$

Denote by P_{B} and $\tilde{P}_{\mathrm{B}}$ the block error probabilities when the transmitted codeword is distributed according to $g(U_{1:m})$ and $X_{1:N}$, respectively. Then,

$$\begin{aligned}
P_{\mathrm{B}} &= \sum_{x\in\{0,1\}^N} \mathbb{P}(\text{error} \mid x)\mathbb{P}_{g(U_{1:m})}(x) \\
&= \sum_{x\in\{0,1\}^N} \mathbb{P}(\text{error} \mid x)(\mathbb{P}_{g(U_{1:m})}(x) - \mathbb{P}_{X_{1:N}}(x)) + \sum_{x\in\{0,1\}^N} \mathbb{P}(\text{error} \mid x)\mathbb{P}_{X_{1:N}}(x) \\
&\leq \sum_{x\in\{0,1\}^N} \mathbb{P}(\text{error} \mid x) \cdot \max(\mathbb{P}_{g(U_{1:m})}(x) - \mathbb{P}_{X_{1:N}}(x), 0) \\
&\quad + \sum_{x\in\{0,1\}^N} \mathbb{P}(\text{error} \mid x)\mathbb{P}_{X_{1:N}}(x) \\
&< \delta + \tilde{P}_{\mathrm{B}},
\end{aligned} \tag{5.20}$$

where the last inequality uses (5.19) and that $\mathbb{P}(\text{error} \mid x) \leq 1$. The requirement (5.20) is crucial because, in the channel coding part of the scheme, we assume that the transmitted codeword has an i.i.d. Bernoulli(α) distribution.

Let us now describe how to construct maps f and g such that these maps satisfy the desired properties. One possible solution is based on polar codes, and the idea follows closely the scheme described in Section 4.4.2. Given $U_{1:m}$ as input, we put it into the positions indexed by $\mathcal{H}_X$ defined in (4.10), and set the remaining positions according to the "randomized rounding" rule (4.20). Then, we multiply this vector with the matrix G_N, and define $g(U_{1:m})$ as the result of this last operation. Given $X_{1:N}$ as input, we multiply it with the matrix G_N, and extract the positions indexed by $\mathcal{H}_X$. We define $f(X_{1:N})$ as the result of this last operation. It is clear that $U_{1:m} = f(g(U_{1:m}))$, hence the first property of Theorem 5.1 is satisfied. By following the proof of Theorem 3 of [121], we also obtain that the total variation

distance between the distribution of $g(U_{1:m})$ and the i.i.d. Bernoulli(α) distribution is upper bounded by 2^{-N^β} for $\beta < 1/2$. Hence, the second property of Theorem 5.1 is satisfied.

An alternative solution is based on arithmetic coding. Let us start by defining the map g. We partition the interval $[0,1)$ into 2^m sub-intervals of size 2^{-m}. Given $U_{1:m}$ as input, we interpret this sequence as the integer $K \in \{0, \cdots, 2^m - 1\}$, and we pick a point $\bar{U}$ uniformly at random in the sub-interval $[K2^{-m}, (K+1)2^{-m})$. The output sequence $g(U_{1:m})$ is obtained from $\bar{U}$ as follows. Given an interval $I = [i_{\text{start}}, i_{\text{end}})$, we partition it into the sub-intervals $I_1 = [i_{\text{start}}, i_{\text{start}} + \alpha(i_{\text{end}} - i_{\text{start}}))$ and $I_2 = [i_{\text{start}} + \alpha(i_{\text{end}} - i_{\text{start}}), i_{\text{end}})$. Note that $|I_1| = \alpha|I|$ and $|I_2| = (1-\alpha)|I|$, where $|I|$, $|I_1|$, and $|I_2|$ denote the sizes of I, I_1, and I_2, respectively. We initialize I to be the interval $[0,1)$. If $\bar{U} \in I_1$, then we output a 1 and redefine I to be I_1; otherwise, we output a 0 and redefine I to be I_2. By repeating this procedure N times, we obtain the sequence $g(U_{1:m})$.

Let us define the map f. Given $X_{1:N}$ as input, we evaluate the interval I according to the following iterative procedure. We initialize I to be the interval $[0,1)$. If the input is 1, we redefine I to be I_1; otherwise, we redefine I to be I_2. As the input sequence $X_{1:N}$ has length N, we repeat this operation N times. Then, we pick a point $\bar{X}$ uniformly at random in the resulting interval I. Let K be such that $\bar{X} \in [K2^{-m}, (K+1)2^{-m})$. We define $f(X_{1:N})$ to be the sequence associated to the integer K.

Let us prove that the maps f and g defined above satisfy the desired properties. As $U_{1:m}$ is a sequence of m i.i.d. uniform random variables, the point $\bar{U}$ is uniformly distributed in $[0,1)$. Hence, the sequence $g(U_{1:m})$ obtained with the aforementioned procedure is exactly a sequence of N i.i.d. Bernoulli(α) random variables. As a result, the second property of Theorem 5.1 holds. Let us prove that also the first property holds. Given a sequence $g(u) \in \{0,1\}^N$, denote by $I(g(u))$ the interval associated with it. Then, for any $\epsilon > 0$,

$$\begin{aligned}\mathbb{P}(f(g(U_{1:m})) \neq U_{1:m}) &= \sum_{u \in \{0,1\}^m : f(g(u)) \neq u} \mathbb{P}_{g(U_{1:m})}(g(u)) \\ &= \sum_{\substack{u \in \{0,1\}^m : f(g(u)) \neq u \\ |I(g(u))| > 2^{-N(h_2(\alpha)-\epsilon)}}} \mathbb{P}_{g(U_{1:m})}(g(u)) + \sum_{\substack{u \in \{0,1\}^m : f(g(u)) \neq u \\ |I(g(u))| \leq 2^{-N(h_2(\alpha)-\epsilon)}}} \mathbb{P}_{g(U_{1:m})}(g(u)).\end{aligned} \tag{5.21}$$

Note that $|I(g(u))| = \alpha^{w_{\text{H}}(g(u))}(1-\alpha)^{N - w_{\text{H}}(g(u))}$, where w_{H} denotes the Hamming weight. In addition, recall that $g(U_{1:m})$ has an i.i.d. Bernoulli(α) distribution. Hence, if $|I(g(u))| > 2^{-N(h_2(\alpha)-\epsilon)}$, then the sequence $g(u)$ cannot be typical because of its Hamming weight. Consequently, we can make the first term of the RHS of (5.21) arbitrarily small. Furthermore, by observing that $\mathbb{P}_{g(U_{1:m})}(g(u)) = |I(g(u))|$, we have that

$$\sum_{\substack{u \in \{0,1\}^m : f(g(u)) \neq u \\ |I(g(u))| \leq 2^{-N(h_2(\alpha)-\epsilon)}}} \mathbb{P}_{g(U_{1:m})}(g(u)) = \sum_{\substack{u \in \{0,1\}^m : f(g(u)) \neq u \\ |I(g(u))| \leq 2^{-N(h_2(\alpha)-\epsilon)}}} |I(g(u))| \leq 2^m \cdot 2^{-N(h_2(\alpha)-\epsilon)}, \tag{5.22}$$

as there are at most 2^m possible sequences u. Hence, by taking $m = N(h_2(\alpha) - 2\epsilon)$, we can make also the second term of the RHS of (5.21) arbitrarily small. This suffices to prove that the first property of Theorem 5.1 holds.

Another possible solution is based on sparse graph codes, and the idea follows closely the scheme described in Section 5.4. Let $P \in \mathbb{F}_2^{m \times N}$ be the parity-check matrix of, e.g., an SC-LDPC code with sufficiently large degrees. Given $U_{1:m}$, we use it to initialize the values of the check nodes, and run belief-propagation guided decimation. Then, we define $g(U_{1:m})$ as the output of the algorithm. Given $X_{1:N}$, we define $f(X_{1:N})$ as $X_{1:N}P^T$. As pointed out in Section 5.4, $U_{1:m}$ differs from $f(g(U_{1:m}))$ in a vanishing fraction of positions, but we can cope with this issue by pre-coding $U_{1:m}$ with a negligible loss in rate. Note that this solution works well in practice, but, to the best of our knowledge, it is not provable.

Note that the second property of Theorem 5.1 is rather stringent. In fact, a weaker condition is sufficient, provided that the transmitter and the receiver have shared randomness. Let us describe this weaker condition in detail. Given a binary sequence $x_{1:N} \in \mathbb{F}_2^N$, let $\tau(x_{1:N})$ denote its type, i.e., the number of 1s contained in the sequence. Then, for Theorem 5.1 to hold, rather than requiring that the distributions of $g(U_{1:m})$ and $X_{1:N}$ are roughly the same, it suffices that the distributions of $\tau(g(U_{1:m}))$ and $\tau(X_{1:N})$ are roughly the same and that a permuted version of $g(U_{1:m})$, call it $\pi(g(U_{1:m}))$, is transmitted over the channel. We require shared randomness, as the random permutation π needs to be shared between the transmitted and the receiver. These concepts are formalized by the following proposition, whose proof immediately follows.

Proposition 5.6. *Denote by $X_{1:N}$ a sequence of N i.i.d. Bernoulli(α) random variables for some $\alpha \in [0,1]$. Let $g(U_{1:m})$ be such that the total variation distance between the distribution of $\tau(g(U_{1:m}))$ and the distribution of $\tau(X_{1:N})$ is at most δ. Let $\pi : [N] \to [N]$ be a random permutation. Then, the total variation distance between the distribution of $\pi(g(U_{1:m}))$ and the distribution of $X_{1:N}$ is at most δ.*

Proof. Note that the type of a sequence is equal to the type of any permutation of such a sequence. By using this fact and the definition of type τ, we have that

$$\begin{aligned}\frac{1}{2}\sum_{x\in\{0,1\}^N} |\mathbb{P}_{\pi(g(U_{1:m}))}(x) - \mathbb{P}_{X_{1:N}}(x)| &= \frac{1}{2}\sum_{i=0}^{N}\sum_{\substack{x\in\{0,1\}^N\\ \tau(x)=i}} |\mathbb{P}_{\pi(g(U_{1:m}))}(x) - \mathbb{P}_{X_{1:N}}(x)| \\ &= \frac{1}{2}\sum_{i=0}^{N} |\mathbb{P}_{\tau(\pi(g(U_{1:m})))}(i) - \mathbb{P}_{\tau(X_{1:N})}(i)| \\ &= \frac{1}{2}\sum_{i=0}^{N} |\mathbb{P}_{\tau(g(U_{1:m}))}(i) - \mathbb{P}_{\tau(X_{1:N})}(i)| \end{aligned} \tag{5.23}$$

On the one hand, the LHS of (5.23) represents the total variation distance between the distribution of $\pi(g(U_{1:m}))$ and the distribution of $X_{1:N}$. On the other hand, the RHS of (5.23) represents the total variation distance between the distribution of $\tau(g(U_{1:m}))$ and the distribution of $\tau(X_{1:N})$. Hence, the claim readily follows. □

As a result, by simply using an extra shared random permutation, if $g(U_{1:m})$ and $X_{1:N}$ have roughly the same type, then we fulfill the rather stringent second property of Theorem 5.1.

In summary, in the source coding part of the scheme, we are given as input the vector $U_{1:m}$ with i.i.d. uniform distribution, and we generate the codeword $g(U_{1:m})$ that is transmitted over the channel. The distribution $g(U_{1:m})$ is close in total variation distance to an i.i.d. Bernoulli(α) distribution. Hence, by paying a negligible price in terms of the error probability, we can assume that the transmitted codeword is the sequence $X_{1:N}$ with i.i.d. Bernoulli(α) distribution. The *channel coding* part of the scheme consists in transmitting reliably $X_{1:N}$ over the channel. A similar problem has been considered in Section 5.2.3. There, we have proved that $X_{1:N}$ can be reconstructed with high probability, given the channel output $Y_{1:N}$ and $NH(X \mid Y)$ additional bits of information. Recall that $h : \{0,1\}^N \to \{0,1\}^\ell$, with ℓ roughly equal to $NH(X \mid Y)$, hence the mapping $h(X_{1:N})$ provides these extra $NH(X \mid Y)$ bits. This means that, if we were able to share the vector $h(X_{1:N})$ between the transmitter and the receiver, we would be done. However, $h(X_{1:N})$ obviously depends on $X_{1:N}$, hence on the information vector $U_{1:m}$. Thus, it is not immediately clear how to share this vector.

To solve this issue, we draw inspiration from the "chaining" construction introduced in [48, 122] and used in Chapter 4 of this thesis to devise polar coding schemes for the broadcast channel. We consider the transmission of k blocks of information, and use a part of the current block to store the parity-check vector of the previous block. More specifically, in block 1, we fill $U_{1:m}$ with information bits, compute $X_{1:N} = g(U_{1:m})$, and $h(X_{1:N})$. In block j ($j \in \{2, \cdots, k-1\}$), we fill $U_{1:m}$ with the vector $h(X_{1:N})$ of the previous block, and store the information bits in the remaining positions. Note that the vector $h(X_{1:N})$ has roughly size $NH(X \mid Y)$, and recall that m is close to $Nh_2(\alpha)$. Hence, the transmission rate is approximately $h_2(\alpha) - H(X \mid Y) = C(W)$. Then, we compute $X_{1:N} = g(U_{1:m})$ and $h(X_{1:N})$. In block k, we transmit only the vector $h(X_{1:N})$ of the previous block at rate $C_{\mathrm{s}}(W)$, by using a code that achieves the symmetric capacity of W (see Section 5.2.2 for polar coding and sparse graph coding schemes that achieve such a goal). Note that, in the last block, we suffer a rate loss, as we are limited to a rate of $C_{\mathrm{s}}(W) < C(W)$. However, this rate loss decays as $1/k$ and, by choosing k large, we achieve a rate arbitrarily close to $C(W)$.

At the receiver, we perform the decoding "backwards", by starting with block k and ending with block 1. More specifically, block k can be easily decoded, as the underlying code achieves the symmetric capacity of the channel. For block j ($j \in \{k-1, \cdots, 1\}$), the decoder can recover $X_{1:N}$ by using the channel output and $h(X_{1:N})$. Indeed, this last vector is obtained from the next block that is already decoded. Finally, from $X_{1:N}$ we deduce $U_{1:m} = f(X_{1:N})$.

Let us now describe how to construct a map h that fulfills the desired property. One possible solution is based on sparse graph codes. Let $P \in \mathbb{F}_2^{\ell \times N}$ be the parity-check matrix of an SC-LDPC code with sufficiently large degrees. Then, we set $h(X_{1:N}) = X_{1:N}P^T$, and conclude that the third property of Theorem 5.1 holds by following the argument of Section 5.2.3. An alternative solution is based on polar coding techniques. We multiply $X_{1:N}$ with the polarizing matrix G_N and extract the positions indexed by the complement of the set $\mathcal{L}_{X|Y}$ defined in (4.16). Then, we define $h(X_{1:N})$ as the result of this last operation. Furthermore, by ap-

plying (4.18), we easily verify that the vector $h(X_{1:N})$ has the correct size, i.e., roughly $NH(X \mid Y)$. Furthermore, by definition, the set $\mathcal{L}_{X|Y}$ contains the positions that are approximately a deterministic function of the previously decoded bits and the output. Hence, we can reconstruct $X_{1:N}$ with high probability given $Y_{1:N}$ and $h(X_{1:N})$, and the required property of h holds.

Encoding. Let $\mathcal{C}_{\mathrm{s}}$ be a code that achieves the symmetric capacity of the channel W and denote by $\mathcal{E}_{\mathrm{s}}$ its encoder. We consider the transmission of k blocks and encode them in order, starting from block 1 and ending with block k.

In block 1, we place the information into $u_{1:m}^{(1)}$, compute the codeword $x_{1:N}^{(1)} = g(u_{1:m}^{(1)})$ and the vector $s_{1:\ell}^{(1)} = h(x_{1:N}^{(1)})$. The codeword $x_{1:N}^{(1)}$ is transmitted over the channel W and the vector $s_{1:\ell}^{(1)}$ is stored into the next block.

In block j ($j \in \{2, \cdots, k-1\}$), we place $s_{1:\ell}^{(j-1)}$ and $NC(W)$ information bits into $u_{1:m}^{(j)}$. Note that this is possible, as $\ell \approx NH(X \mid Y)$, $m \approx Nh_2(\alpha)$, and $H(X \mid Y) + C(W) = h_2(\alpha)$. Then, we compute the codeword $x_{1:N}^{(j)} = g(u_{1:m}^{(j)})$, and the vector $s_{1:\ell}^{(j)} = h(x_{1:N}^{(j)})$. Once again, the codeword $x_{1:N}^{(j)}$ is transmitted over the channel W and the vector $s_{1:\ell}^{(j)}$ is stored into the next block.

In block k, we place $s_{1:\ell}^{(k-1)}$ into $u_{1:\ell}^{(k)}$. Then, we map $u_{1:\ell}^{(k)}$ into the codeword $x_{1:N'}^{(k)}$ via the encoder $\mathcal{E}_{\mathrm{s}}$. Note that N' is an integer roughly equal to $\ell / C_{\mathrm{s}}(W)$, as the code $\mathcal{C}_{\mathrm{s}}$ has rate close to $C_{\mathrm{s}}(W)$.

The overall rate of communication is given by

$$R = \frac{Nh_2(\alpha) + N(k-2)C(W)}{N(k-1) + NH(X \mid Y)/C_{\mathrm{s}}(W)}, \tag{5.24}$$

that, as k goes large, tends to the required rate $C(W)$.

Decoding. Denote by $\mathcal{D}_{\mathrm{s}}$ the decoder of the code $\mathcal{C}_{\mathrm{s}}$ and by $\mathcal{D}$ the decoder that recovers the codeword $x_{1:N}$ given the channel output $y_{1:N}$ and the vector $h(x_{1:N})$. The decoding process begins after all the k blocks have been received, and it operates "backwards", starting from block k and ending with block 1.

In block k, the decoder $\mathcal{D}_{\mathrm{s}}$ accepts as input the received message $y_{1:N'}^{(k)}$. The output is the estimate $\hat{u}_{1:\ell}^{(k)}$ on the payload $u_{1:\ell}^{(k)}$ of block k. This immediately yields the estimate $\hat{s}_{1:\ell}^{(k-1)}$ on the vector $s_{1:\ell}^{(k-1)}$ of block $k-1$.

In block j ($j \in \{k-1, \cdots, 2\}$), the decoder $\mathcal{D}$ accepts as inputs the received message $y_{1:N}^{(j)}$ and the previously obtained estimate $\hat{s}_{1:\ell}^{(j)}$. The output is the estimate $\hat{x}_{1:N}^{(j)}$ on the codeword $x_{1:N}^{(j)}$ of block j. Then, we compute $\hat{u}_{1:m}^{(j)} = f(\hat{x}_{1:N}^{(j)})$, which yields an estimate on the information bits transmitted in block j and on the vector $s_{1:\ell}^{(j-1)}$ of block $j-1$.

For block 1 the decoding process is the same as that for block j ($j \in \{k-1, \cdots, 2\}$). The only difference consists in the fact that $\hat{u}_{1:m}^{(1)}$ contains solely an estimate on information bits.

The situation is schematically represented in Figure 5.3.

Performance. There are four possible types of errors.

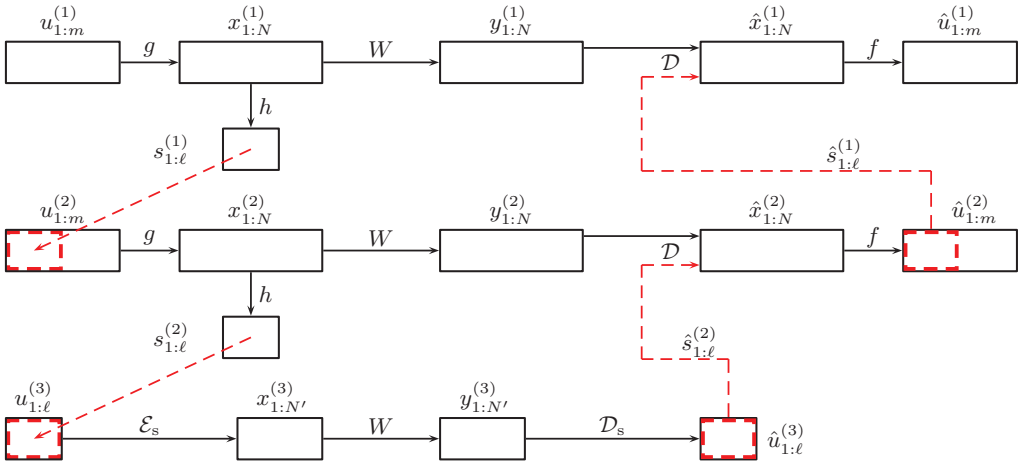


Figure 5.3 – Coding over asymmetric channels via the chaining construction. We consider the transmission of $k = 3$ blocks, and store the vector of parity checks of block $j-1$ into block j ($j \in \{2, 3\}$).

1. In block j ($j \in \{1, \cdots, k-1\}$), given that $\hat{x}^{(j)}_{1:N} = x^{(j)}_{1:N}$, we might have that $\hat{u}^{(j)}_{1:m} \neq u^{(j)}_{1:m}$.
2. In block j ($j \in \{1, \cdots, k-1\}$), the encoder might fail to produce a codeword $x^{(j)}_{1:N}$ with the correct distribution (namely, with roughly $N\alpha$ 1s).
3. In block j ($j \in \{1, \cdots, k-1\}$), we might have that $\hat{x}^{(j)}_{1:N} \neq x^{(j)}_{1:N}$.
4. In block k, we might have that $\hat{x}^{(k)}_{1:N} \neq x^{(k)}_{1:N}$.

By hypothesis, the map f fulfills the first property stated in Theorem 5.1. Hence, by the union bound, the probability that the first error event takes place is at most $(k-1)\delta$. Similarly, the map g fulfills the second property stated in Theorem 5.1. Hence, by the union bound, the probability that the second error event takes place is at most $(k-1)\delta$. Furthermore, the map h fulfills the third property stated in Theorem 5.1. Hence, by the union bound, the probability that the third error event takes place is at most $(k-1)\delta$. As the code $\mathcal{C}_{\mathrm{s}}$ achieves the symmetric capacity of the channel W, the last event occurs with probability at most δ. As a result, the error probability of the proposed scheme is upper bounded by $3k\delta$. Recall that k is large but fixed (it only depends on the rate we want to achieve), hence, by choosing δ sufficiently small, the proof of Theorem 5.1 is complete.

5.6 Performance Comparison between the Three Paradigms

5.6.1 Error Probability

First, consider *Gallager's mapping*. Recall that in Section 5.3 we describe two schemes: one based on a single non-binary code $\mathcal{C}$, and the other based on t binary

codes $\{\mathcal{C}_j\}_{j\in[t]}$. Let W be the transmission channel and let W' be defined as in (5.9). Then, for the scheme based on a single non-binary code, the error probability is the same as that of the transmission of $\mathcal{C}$ over W'. For $j \in [t]$, let W''_j be defined as in (5.12). Then, for the scheme based on t binary codes, the error probability is upper bounded by the sum over j of the error probabilities of the transmission of $\mathcal{C}_j$ over W''_j. This means that we need to multiply the error probability by a factor of t.

Second, consider the *integrated scheme*. In [121] the authors provide a comparison between the second-order error exponent of Gallager's mapping and of the integrated scheme with polar codes reviewed in Section 4.4.2. In particular, let p be the input distribution induced by Gallager's mapping. Then, if the transmission rate R is sufficiently close[2] to $I(p)$, the integrated scheme achieves a better second-order error exponent than Gallager's mapping.

Third, consider the *chaining construction*. Recall that in Section 5.5 we have divided the transmission in k blocks and performed the decoding "backwards". This method suffers from error propagation, in the sense that an error occurring in block t propagates to all the previous blocks from $t-1$ to 1. Hence, we need to multiply the error probability by a factor of k. Note that such a behavior occured also in Chapter 4. More specifically, in formulae (4.48) and (4.64) there is a factor of k in the expression of the error probability. However, in the case of polar codes, this fact does not influence much the scaling of the error probability. Indeed, for a fixed rate $R < C(W)$, the error probability under successive cancellation decoding scales as $2^{-\sqrt{N}}$ [56], and the number of blocks k is a constant independent of N.

5.6.2 Rate Penalty

First, consider *Gallager's mapping*. In this case, the rate penalty comes from the fact that the distribution p induced by the map might not be exactly equal to the capacity-achieving input distribution p^*. We quantify the rate penalty in Proposition 5.5 in terms of the total variation distance between p and p^*, and of the cardinalities of the input and output alphabets. The rate penalty is also studied in [121] for the special case of binary-input alphabet. Note that the bound obtained in formula (25) of [121] is tighter[3] than our bound of Proposition 5.5, but it is significantly less general as it crucially uses the fact that the input alphabet is binary.

Second, consider the *integrated scheme*. In this case, the rate penalty comes from the fact that we need to pre-code the syndrome vector, as the output of the belief-propagation guided decimation algorithm does not coincide exactly with the given syndrome vector. In Section 5.4, we have observed that the fraction of unfulfilled syndromes tends 0 as N goes large. Hence, also the rate penalty can be made arbitrarily small. However, the rigorous proof of these statements remains an open problem.

Third, consider the *chaining construction*. In this case, the rate penalty comes from the fact that the rate in the last block is $C_{\mathrm{s}}(W) < C(W)$, where $C(W)$ and $C_{\mathrm{s}}(W)$ denote the capacity and the symmetric capacity of the channel W, respec-

[2]The exact condition is $R > I(p) - I(p^*)(I(p^*) - I(p))$, where p^* is the capacity-achieving input distribution.

[3]Let δ be the total variation distance between p and p^*. Then, formula (25) of [121] gives that the rate penalty is $O(\delta^2)$, whereas Proposition 5.5 gives that the rate penalty is $O(\delta \log_2(1/\delta))$.

tively. From formula (5.24), we immediately obtain that the rate penalty is $O(1/k)$, where k is the number of blocks.

5.6.3 Computational Complexity

First, consider *Gallager's mapping*. The computational complexity of this scheme scales as a linear function of the cardinality[4] of the domain of the map. Furthermore, note that the total variation distance between p and p^*, call it δ, scales with the inverse of this cardinality. Hence, the computational complexity scales as $1/\delta$.

Second, consider the *integrated scheme*. This approach has the same computational complexity as the standard channel coding solution for the transmission over a symmetric channel.

Third, consider the *chaining construction*. The computational complexity of this scheme scales as a linear function of the number of blocks k, as there are k blocks to be encoded and decoded.

5.6.4 Universality

We say that a coding scheme achieves capacity *universally* over a class of channels if it achieves the capacity of each channel in the class at the same time. This means that the coding scheme is not tailored to the specific channel, rather it can be used for the transmission over any channel in the class. SC-LDPC codes universally achieve capacity over the class of B-DMCs [36]. Polar codes, on the contrary, are not universal, as the sets defined in (4.15) and (4.16) depend on the transmission channel. Therefore, several "polar-like" schemes have been developed to solve this issue [48, 49].

First, consider *Gallager's mapping*. This scheme is not universal, as different transmission channels require different capacity-achieving distributions, hence different mappings. On the contrary, the *integrated scheme* and the *chaining construction* are universal, provided that the underlying component codes are universal (e.g., we use either SC-LDPC codes or the "polar-like" schemes described in [48, 49]).

5.6.5 Delay

Gallager's mapping and the *integrated scheme* have the same delay as the standard channel coding solution for the transmission over a symmetric channel. The *chaining construction*, on the contrary, suffers a delay that scales as a linear function of the number of blocks k. Indeed, as described in Section 5.5, the decoding process starts after that all the k blocks have been received.

5.6.6 Common Randomness

First, consider *Gallager's mapping*. This scheme does not require common randomness, as the maps f and g defined in Section 5.3 are deterministic.

Second, consider the *integrated scheme*. This approach requires common randomness both in the version based on polar codes and in the version based on

[4] Recall from Section 5.3.2 that, for the scheme based on a single non-binary code, this cardinality is $|\mathcal{V}|$, and that, for the scheme based on t binary codes, this cardinality is $|\mathcal{U}| = 2^t$.

sparse graph codes. More specifically, in the polar version reviewed in Section 4.4.2, we fill the positions in $\mathcal{F}_{\mathrm{r}}$ with a sequence chosen uniformly at random and shared between the transmitted and the receiver. Furthermore, we encode the positions in $\mathcal{F}_{\mathrm{d}}$ via the random map defined in (4.20) that also needs to be shared. In the sparse graph version described in Section 5.4, the syndrome vector $S^{(2)}_{1:NH(X|Y)}$ is chosen uniformly at random and shared between the transmitter and the receiver.

Third, consider the *chaining construction*. By separating the source coding and channel coding parts of the scheme, this approach does not require common randomness, hence it can be interpreted as a derandomized version of the integrated scheme. This establishes another connection between information theory and the theory of derandomizing algorithms. Several applications of derandomization to coding theory can be found in [184], i.e., information-theoretically secure schemes for the wiretap channel, nearly optimal explicit measurement schemes for combinatorial group testing, design of ensembles of capacity-achieving codes, and construction of codes arbitrarily close to the Gilbert-Varshamov bound. Furthermore, the link between polarization and randomness extraction is investigated in [185], where applications to the Slepian-Wolf problem and to secret key generation are provided.

Assume that we want to substitute the stringent condition on the distance between the distributions of $g(U_{1:m})$ and $X_{1:N}$ with the relaxed condition on the distance between the distributions of their types. Then, as detailed in Proposition 5.6, the transmitter and the receiver need to share k random permutations, where k is the number of blocks used for the transmission. We will now show that no shared randomness is in fact necessary.

As a starting point, recall that the error probability under the stringent condition with no random permutation is the same as the error probability under the relaxed condition with random permutations. Furthermore, this probability is upper bounded by $3k\delta$.

The error probability is an average over all channel realizations and all permutations. Hence, for any $\gamma > 0$, at least a fraction $1-\gamma$ of the permutations have an error probability of at most $3k\delta/\gamma$. By picking $\gamma = \sqrt{3k\delta}$, we have that, with probability at least $1-\sqrt{3k\delta}$, a randomly chosen permutation has an error probability of at most $\sqrt{3k\delta}$. Hence, no shared randomness is needed, as a fixed set of k permutations will work with high probability.

5.7 Appendix

5.7.1 Proof of Propositions in Section 5.2.2

Proof of Proposition 5.1. By definition, we have that

$$\begin{aligned}\mathsf{a}^{+}(y)\Delta y &\approx \int_{t\in L^{-1}([y,y+\Delta y])} W(t \mid 1)\, dt = \int_{t\in L^{-1}([y,y+\Delta y])} e^{L(t)} W(t \mid -1)\, dt \\ &\approx e^{y} \int_{t\in L^{-1}([y,y+\Delta y])} W(t \mid -1)\, dt = e^{y}\mathsf{a}^{-}(y)\Delta y,\end{aligned}$$

where L^{-1} is the inverse of the log-likelihood ratio defined in (5.1). By taking $\Delta y \to 0$, we obtain that

$$\mathsf{a}^+(y) = e^y \mathsf{a}^-(y). \tag{5.25}$$

With the change of variable $y \to -y$, we also obtain that

$$\mathsf{a}^-(-y) = e^y \mathsf{a}^+(-y). \tag{5.26}$$

As a result, condition (5.2) is fulfilled for the L-density $\mathsf{a}^{\mathrm{s}}(y)$ defined in (5.4) and the statement follows. □

Proof of Proposition 5.2. Since the log-likelihood ratio constitutes a sufficient statistic, two B-DMCs are equivalent if they have the same L-densities given that $X = \pm 1$ is transmitted. As a representative for the equivalence class, we can take

$$\begin{aligned} W(y \mid 1) &= \mathsf{a}^+(y), \\ W(y \mid -1) &= \mathsf{a}^-(y). \end{aligned} \tag{5.27}$$

By definition of log-likelihood ratio and by using (5.25), we have

$$L(y) = \ln \frac{W(y \mid 1)}{W(y \mid -1)} = \ln \frac{\mathsf{a}^+(y)}{\mathsf{a}^-(y)} = y.$$

Therefore,

$$\lim_{\Delta y \to 0} \frac{\mathbb{P}(L(Y) \in [y, y + \Delta y] \mid X = \pm 1)}{\Delta y} = \mathsf{a}^{\pm}(y),$$

which means that (5.27) is a valid choice.

Let X be uniformly distributed. Then, after some calculations we have that

$$\begin{aligned} C_{\mathrm{s}}(W) = H(Y) - H(Y \mid X) = \frac{1}{2} \int W(y \mid 1) \log_2 \frac{2W(y \mid 1)}{W(y \mid 1) + W(y \mid -1)} \, dy \\ + \frac{1}{2} \int W(y \mid -1) \log_2 \frac{2W(y \mid -1)}{W(y \mid 1) + W(y \mid -1)} \, dy. \end{aligned} \tag{5.28}$$

By applying (5.27) and (5.25), the first integral simplifies to

$$\frac{1}{2} \int \mathsf{a}^+(y) \left(1 - \log_2(1 + e^{-y})\right) \, dy. \tag{5.29}$$

By applying (5.27), doing the change of variables $y \to -y$ and using (5.26), the second integral simplifies to

$$\frac{1}{2} \int \mathsf{a}^-(-y) \left(1 - \log_2(1 + e^{-y})\right) \, dy. \tag{5.30}$$

By combining (5.28), (5.29), and (5.30), the result follows.

□

5.7.2 Proof of Propositions in Section 5.2.3

Proof of Proposition 5.3. By definition, we have that

$$\begin{aligned}\bar{\alpha}\mathsf{a}_{\mathrm{p}}^{+}(y)\Delta y &\approx \int_{t\in L_{\mathrm{p}}^{-1}([y,y+\Delta y])} \bar{\alpha}W(t\mid 1)\,dt = \int_{t\in L_{\mathrm{p}}^{-1}([y,y+\Delta y])} \alpha e^{L_{\mathrm{p}}(t)}W(t\mid -1)\,dt\\ &\approx e^{y}\int_{t\in L_{\mathrm{p}}^{-1}([y,y+\Delta y])} \alpha W(t\mid -1)\,dt = e^{y}\alpha\mathsf{a}_{\mathrm{p}}^{-}(y)\Delta y,\end{aligned}$$

where L_{p}^{-1} is the inverse of L_{p} defined in (5.6). By taking $\Delta y \to 0$, we obtain that

$$\bar{\alpha}\mathsf{a}_{\mathrm{p}}^{+}(y) = e^{y}\alpha\mathsf{a}_{\mathrm{p}}^{-}(y). \tag{5.31}$$

With the change of variable $y \to -y$, we also obtain that

$$\alpha\mathsf{a}_{\mathrm{p}}^{-}(-y) = \bar{\alpha}e^{y}\mathsf{a}_{\mathrm{p}}^{+}(-y). \tag{5.32}$$

As a result, condition (5.2) is fulfilled for $\mathsf{a}_{\mathrm{p}}^{\mathrm{s}}(y)$ and the statement follows. □

Proof of Proposition 5.4. Since the log-likelihood ratio constitutes a sufficient statistic, two B-DMCs with non-uniform input distributions are equivalent if they have the same densities of the log-posterior ratio given that $X = \pm 1$ is transmitted. As a representative for the equivalence class, we can take

$$\begin{aligned}W(y\mid 1) &= \mathsf{a}_{\mathrm{p}}^{+}(y),\\ W(y\mid -1) &= \mathsf{a}_{\mathrm{p}}^{-}(y).\end{aligned} \tag{5.33}$$

By definition of log-posterior ratio and by using (5.31), we have

$$L_{\mathrm{p}}(y) = \ln\frac{p_{X|Y}(1\mid y)}{p_{X|Y}(-1\mid y)} = \ln\frac{\bar{\alpha}\mathsf{a}_{\mathrm{p}}^{+}(y)}{\alpha\mathsf{a}_{\mathrm{p}}^{-}(y)} = y.$$

Therefore,

$$\lim_{\Delta y\to 0}\frac{\mathbb{P}(L_{\mathrm{p}}(Y)\in[y,y+\Delta y]\mid X=\pm 1)}{\Delta y} = \mathsf{a}_{\mathrm{p}}^{\pm}(y),$$

which means that (5.33) is a valid choice.

Let $X \in \{-1,1\}$ be such that $\mathbb{P}(X=-1)=\alpha$. Then, after some calculations we have that

$$\begin{aligned}H(X\mid Y) = &-\int \bar{\alpha}W(y\mid 1)\log_2\frac{\bar{\alpha}W(y\mid 1)}{\bar{\alpha}W(y\mid 1)+\alpha W(y\mid -1)}dy\\ &-\int \alpha W(y\mid -1)\log_2\frac{\alpha W(y\mid -1)}{\bar{\alpha}W(y\mid 1)+\alpha W(y\mid -1)}dy.\end{aligned} \tag{5.34}$$

By applying (5.33) and (5.31), the first integral simplifies to

$$\int \bar{\alpha}\mathsf{a}_{\mathrm{p}}^{+}(y)\left(1-\log_2(1+e^{-y})\right)\,dy. \tag{5.35}$$

By applying (5.33), doing the change of variables $y \to -y$ and using (5.32), the second integral simplifies to

$$\int \alpha\mathsf{a}_{\mathrm{p}}^{-}(-y)\left(1-\log_2(1+e^{-y})\right)\,dy. \tag{5.36}$$

By combining (5.34), (5.35), and (5.36), the result follows.

□

5.7.3 Proof of Proposition 5.5

Before starting with the proof of the proposition, let us state the following useful result [186] that is a refinement of Lemma 2.7 of [187].

Lemma 5.1. *Consider two distributions p and p^* over the alphabet $\mathcal{X}$ such that their total variation distance is equal to δ, i.e., $\frac{1}{2}\sum_{x\in\mathcal{X}}|p^*(x)-p(x)|=\delta$. Take $X\sim p$ and $X^*\sim p^*$. Then,*

$$|H(X^*)-H(X)|\leq \delta\log_2(|\mathcal{X}|-1)+h_2(\delta)<\delta\log_2|\mathcal{X}|+h_2(\delta). \tag{5.37}$$

Proof of Proposition 5.5. Let $X\sim p$, $X^*\sim p^*$ and denote by $Y\sim p_Y$ and $Y^*\sim p_Y^*$ the outputs of the channel when the input is X and X^*, respectively. Denote by $W(y\mid x)$ the probability distribution associated with the channel W. In order to prove (5.14), we write

$$|I(p^*)-I(p)|\leq |H(Y^*)-H(Y)|+|H(Y^*\mid X^*)-H(Y\mid X)|, \tag{5.38}$$

and we bound both terms as functions of δ and $|\mathcal{Y}|$. For the first term, observe that

$$\frac{1}{2}\sum_{y\in\mathcal{Y}}|p_Y^*(y)-p_Y(y)|\leq\frac{1}{2}\sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}}W(y\mid x)|p^*(x)-p(x)|<\delta, \tag{5.39}$$

where it is used the fact that $\sum_{y\in\mathcal{Y}}W(y\mid x)=1$ for any $x\in\mathcal{X}$. Then, by using Lemma 5.1 and the fact that $h_2(\delta)$ is increasing for any $\delta\in(0,1/2)$, we obtain that

$$|H(Y^*)-H(Y)|<\delta\log_2|\mathcal{Y}|+h_2(\delta). \tag{5.40}$$

For the second term, observe that the conditional distribution of Y^* given $X^*=x$ and the conditional distribution of Y given $X=x$ are both equal to $W(y\mid x)$. Therefore,

$$H(Y\mid X=x)=H(Y^*\mid X^*=x)\leq\log_2|\mathcal{Y}|.$$

Consequently,

$$|H(Y^*\mid X^*)-H(Y\mid X)|\leq\sum_{x\in\mathcal{X}}|p^*(x)-p(x)|H(Y\mid X=x)<2\delta\log_2|\mathcal{Y}|. \tag{5.41}$$

By combining (5.38) with (5.40) and (5.41), we obtain the desired result.

In order to prove (5.15), we write

$$|I(p^*)-I(p)|\leq|H(X^*)-H(X)|+|H(X^*\mid Y^*)-H(X\mid Y)|, \tag{5.42}$$

and we bound both terms with functions of δ and $|\mathcal{X}|$. The first term is easily bounded by using Lemma 5.1 and the fact that $h_2(\delta)$ is increasing for any $\delta\in(0,1/2)$,

$$|H(X^*)-H(X)|<\delta\log_2|\mathcal{X}|+h_2(\delta). \tag{5.43}$$

For the second term, consider the conditional distribution of X^* given $Y^*=y$, i.e., $p^*_{X|Y}(x\mid y)=p^*(x)W(y\mid x)/p_Y^*(y)$, and the conditional distribution of X given

$Y = y$, i.e., $p_{X|Y}(x \mid y) = p(x)W(y \mid x)/p_Y(y)$. Then,

$$\begin{aligned} |H(X^* \mid Y^*) - H(X \mid Y)| &= |\sum_{y\in\mathcal{Y}} p_Y^*(y)H(X^*|Y^* = y) - p_Y(y)H(X|Y = y)| \\ &\le |\sum_{y\in\mathcal{Y}} p_Y^*(y)H(X^*|Y^* = y) - p_Y(y)H(X^*|Y^* = y)| \\ &\quad + |\sum_{y\in\mathcal{Y}} p_Y(y)H(X^*|Y^* = y) - p_Y(y)H(X|Y = y)|. \end{aligned} \tag{5.44}$$

In order to bound the first term of (5.44), observe that $H(X^*|Y^* = y) \le \log_2 |\mathcal{X}|$ for any $y \in \mathcal{Y}$. Therefore, by using (5.39), we obtain

$$|\sum_{y\in\mathcal{Y}} p_Y^*(y)H(X^*|Y^* = y) - p_Y(y)H(X^*|Y^* = y)| < 2\delta \log_2 |\mathcal{X}|. \tag{5.45}$$

For the second term of (5.44), let us denote by $d(y)$ the total variation distance between $p^*_{X|Y}(x \mid y)$ and $p_{X|Y}(x \mid y)$, namely,

$$d(y) = \frac{1}{2}\sum_{x\in\mathcal{X}} |p^*_{X|Y}(x \mid y) - p_{X|Y}(x \mid y)|.$$

Then, by Lemma 5.1,

$$|H(X^*|Y^* = y) - H(X|Y = y)| < d(y) \log_2 |\mathcal{X}| + h_2(d(y)),$$

which implies that

$$\begin{aligned} |\sum_{y\in\mathcal{Y}} p_Y(y)H(X^*|Y^* = y) - p_Y(y)H(X|Y = y)| &< \log_2 |\mathcal{X}| \sum_{y\in\mathcal{Y}} p_Y(y)d(y) \\ &\quad + \sum_{y\in\mathcal{Y}} p_Y(y)h_2(d(y)). \end{aligned} \tag{5.46}$$

Now, let us focus on the quantity $\sum_{y\in\mathcal{Y}} p_Y(y)d(y)$:

$$\begin{aligned} \sum_{y\in\mathcal{Y}} p_Y(y)d(y) &= \sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}} |p_Y(y)p^*_{X|Y}(x|y) - p_Y(y)p_{X|Y}(x|y)| \\ &\le \sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}} |p_Y(y)p^*_{X|Y}(x|y) - p^*_Y(y)p^*_{X|Y}(x|y)| \\ &\quad + \sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}} |p^*_Y(y)p^*_{X|Y}(x|y) - p_Y(y)p_{X|Y}(x|y)| \\ &= \sum_{y\in\mathcal{Y}} |p^*_Y(y) - p_Y(y)| \sum_{x\in\mathcal{X}} p^*_{X|Y}(x|y) + \sum_{x\in\mathcal{X}} |p^*(x) - p(x)| \sum_{y\in\mathcal{Y}} W(y \mid x) < 4\delta. \end{aligned}$$

Observe that $h_2(t)$ is concave for any $t \in (0, 1)$ and increasing for $t \le 1/2$. Then, as $\delta < 1/8$,

$$\sum_{y\in\mathcal{Y}} p(y)h_2(d(y)) \le h_2(\sum_{y\in\mathcal{Y}} p(y)d(y)) < h_2(4\delta).$$

By combining (5.42) with (5.43), (5.44), (5.45), and (5.46), the result follows. □

Interlude – from Polar to Reed-Muller Codes

6

Non fare frasi di una sola parola.
Eliminale.

Don't write one-word sentences. Ever.

This chapter[1] connects the seemingly different families of polar codes and Reed-Muller codes, thus serving as an interlude before the last part of this thesis. More specifically, we present an **interpolation method** between the polar code of block length N and rate R and a Reed-Muller code of the same block length and rate. The result is relevant in practice because the codes from this new interpolating family **boost** the **finite-length performance of polar codes** under low-complexity decoding algorithms, such as belief propagation and the successive cancellation list decoder proposed in [55].

In Section 6.1, after pointing out similarities and differences between the polar and the Reed-Muller construction, we describe explicitly the interpolating family $\{\mathcal{C}_\alpha\}$ for the special case of the BEC. In Section 6.2, by starting from the analysis of the two extreme cases of MAP and SC decoding, we show how to improve significantly the finite-length performance of polar codes under practical decoding algorithms (e.g., BP and SCL) by using codes from the family $\{\mathcal{C}_\alpha\}$. In Section 6.3, we generalize the interpolation method to the transmission over any BMS channel and, for a case study, we present the simulation results for the binary-input Gaussian channel.

6.1 Interpolation Method for the BEC

Let us first recall some definitions and concepts that we previously introduced in Section 1.3 and 1.4. Let $n \in \mathbb{N}$ and $N = 2^n$. Consider the $N \times N$ matrix $F^{\otimes n}$, given by n-th Kronecker power of $F = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$. Then, the generator matrices of

[1] The material of this chapter is based on joint work with S. H. Hassani and R. Urbanke [188,189].

both polar and Reed-Muller codes are obtained by suitably selecting rows from the matrix $F^{\otimes n} = (g_1, \cdots, g_N)^T$ (cf. formula (1.13), where B_N is a permutation matrix, hence it simply changes the order of the rows of $F^{\otimes n}$).

On the one hand, the *Reed-Muller rule* consists in choosing the rows of $F^{\otimes n}$ with the largest Hamming weight. Recall that, given n, $v \in \mathbb{N}$, a Reed-Muller code $\mathrm{RM}(n, v)$ is obtained by picking all the rows of $F^{\otimes n}$ with weight at least 2^{n-v}. This code has block length $N = 2^n$, rate $R = \sum_{i=0}^{v} \binom{n}{i}/N$ and minimum distance 2^{n-v}. In general, if we require a Reed-Muller code with fixed block length N and rate R such that NR cannot be written as a sum of binomial coefficients, then we take as generator matrix any subset of NR rows of $F^{\otimes n}$ with the highest Hamming weights. Notice that this criterion is channel-independent, in the sense that it does not rely on the particular channel over which the transmission takes place.

On the other hand, the *polar rule* is channel-specific. Indeed, the N synthetic channels $W_n^{(i)}$, $i \in [N]$, are obtained from N independent copies of the original channel W. The row g_i is associated with $W_n^{(i)}$ and the synthetic channels (hence the rows) with the lowest Bhattacharyya parameters are selected. In general, different channels W yield different choices of rows. Consider the simple case of the transmission over the $\mathrm{BEC}(\varepsilon)$, for $\varepsilon \in (0, 1)$. In order to compute the Bhattacharyya parameter $Z_n^{(i)}$ associated with $W_n^{(i)}$ (hence with g_i), it is easy to check that each 1 in the binary expansion of $i-1$ corresponds to a "+" transform and each 0 to a "−" transform (cf. formula (3.67) in the proof of Lemma 3.5 in Section 3.7.2). Thus,

$$Z_n^{(i)}(\varepsilon) = f_{b_1^{(i)}} \circ f_{b_2^{(i)}} \circ \cdots f_{b_n^{(i)}}(\varepsilon), \tag{6.1}$$

where $f_0(x) = 1 - (1-x)^2$, $f_1(x) = x^2$, $\circ$ denotes function composition, and $b^{(i)} = (b_1^{(i)}, b_2^{(i)}, \cdots, b_n^{(i)})^T$ is the binary expansion of $i-1$ over n bits, $b_1^{(i)}$ being the most significant bit and $b_n^{(i)}$ the least significant bit. In order to construct a code of block length N and rate R, we select the NR rows that minimize the expression (6.1).

The link between the Reed-Muller rule and the polar rule is clarified by the following proposition.

Proposition 6.1. *The polar code of block length N and rate R designed for the transmission over a BEC(ε), when $\varepsilon \to 0$, is a Reed-Muller code.*

Proof. Suppose that the thesis is false, i.e., that we include g_{j^*}, but not g_{i^*}, with $w_{\mathrm{H}}(g_{i^*}) > w_{\mathrm{H}}(g_{j^*})$, where $w_{\mathrm{H}}(\cdot)$ denotes the Hamming weight. Since $w_{\mathrm{H}}(g_i) = 2^{\sum_{k=1}^{n} b_k^{(i)}} = 2^{w_{\mathrm{H}}(b^{(i)})}$ for any $i \in [N]$ from [37, Proposition 17]), then $w_{\mathrm{H}}(b^{(i^*)}) > w_{\mathrm{H}}(b^{(j^*)})$.

From formula (6.1), we deduce that $Z_n^{(i)}(\varepsilon)$ is a polynomial in ε with minimum degree equal to $2^{w_{\mathrm{H}}(b^{(i)})}$. Hence,

$$\lim_{\varepsilon \to 0} \frac{Z_n^{(i^*)}(\varepsilon)}{Z_n^{(j^*)}(\varepsilon)} = 0,$$

which means that there exists $\delta > 0$ such that for all $\varepsilon < \delta$, $Z_n^{(i^*)}(\varepsilon) < Z_n^{(j^*)}(\varepsilon)$.

Consider a polar code designed for the transmission over a $\mathrm{BEC}(\varepsilon)$, with $\varepsilon < \delta$. Then, if this code includes g_{j^*}, it must also include g_{i^*}, which is a contradiction. □

Recall that the transmission takes place over $W = \mathrm{BEC}(\varepsilon)$. Let $\mathcal{C}_\alpha$ be the polar code of block length N and rate R designed for a $\mathrm{BEC}(\alpha\varepsilon)$. When $\alpha = 1$, $\mathcal{C}_\alpha$ reduces to the polar code for the channel W, whereas, when $\alpha \to 0$, $\mathcal{C}_\alpha$ becomes a Reed-Muller code by Proposition 6.1. Now, the codes of the family $\{\mathcal{C}_\alpha\}$ provide an *interpolation method* for passing smoothly from a polar code to a Reed-Muller code of the same rate and block length. Indeed, consider the generator matrices of the codes $\{\mathcal{C}_\alpha\}$ that are obtained by reducing α from 1 to 0. We start from the generator matrix of the polar code, and the successive matrices are obtained by changing one row at a time. In particular, numerical simulations show that the row included in the next code (associated with a smaller α) has a Hamming weight higher than the row removed from the previous code (associated with a higher α). Heuristically, this happens for the following reason. The row indices chosen by $\mathcal{C}_\alpha$ are those that minimize the associated Bhattacharyya parameters $Z_n^{(i)}(\alpha\varepsilon)$ given by (6.1). As $f_1(x) \leq f_0(x)$ for any $x \in [0,1]$, applying f_1 instead of f_0 makes the Bhattacharyya parameter decrease. However, also the order in which the functions are applied is important, since $f_0 \circ f_1(x) \leq f_1 \circ f_0(x)$ for any $x \in [0,1]$: if we fix $w_{\mathrm{H}}(b^{(i)})$, $Z_n^{(i)}$ is minimized by applying first all the functions f_1 and then the functions f_0. Therefore, the goodness of the index i depends both on the number of 1s in its binary expansion $b^{(i)}$ and on the positions of these 1s. Whereas, when designing a Reed-Muller code only $w_{\mathrm{H}}(b^{(i)})$ matters and, for α small enough, $\mathcal{C}_\alpha$ tends to a Reed-Muller code. As a result, as α goes from 1 to 0, the value of $Z_n^{(i)}(\alpha\varepsilon)$ depends more and more on $w_{\mathrm{H}}(b^{(i)})$ rather than on the position of the 1s in $b^{(i)}$.

6.2 Finite-Length Performance Improvement

The focus of this section is on the performance of the codes in the family $\{\mathcal{C}_\alpha\}$ when the transmission takes place over the $\mathrm{BEC}(\varepsilon)$. We start by considering the MAP decoder, then we move to the SC decoder originally introduced by Arıkan. By taking into account low-complexity suboptimal decoding schemes that outperform the original SC algorithm (e.g., SCL and BP), we highlight the advantage of employing codes of the form $\mathcal{C}_\alpha$. The simulation results of this section refer to codes of fixed block length $N = 2^{10}$ and rate $R = 0.5$. The number of Monte Carlo trials is $M = 10^5$.

6.2.1 Motivation: MAP Decoding

It has been observed that, under MAP decoding, picking the rows of $F^{\otimes n}$ according to the Reed-Muller rule significantly improves the performance with respect to the polar choice [53]. Hence, it is interesting to analyze the error probability $P_{\mathrm{B}}^{\mathrm{MAP}}(\alpha, \varepsilon)$ under MAP decoding for the transmission of the code $\mathcal{C}_\alpha$ over the $\mathrm{BEC}(\varepsilon)$. Although MAP decoding is in general an NP-complete task, for the particular case of the BEC it is equivalent to the inversion of a suitable matrix, hence it can be performed in $O(N^3)$.

First of all, fix the value of ε and consider how $P_{\mathrm{B}}^{\mathrm{MAP}}$ varies as a function of α. As it is shown in Figure 6.1 for four distinct values of ε, $P_{\mathrm{B}}^{\mathrm{MAP}}(\alpha, \varepsilon)$ is increasing in α. In short, the proposed interpolation method to pass from the polar code $\mathcal{C}_\alpha\big|_{\alpha=1}$ to a Reed-Muller code $\mathcal{C}_\alpha\big|_{\alpha=0}$ yields a family of codes with *decreasing* MAP error

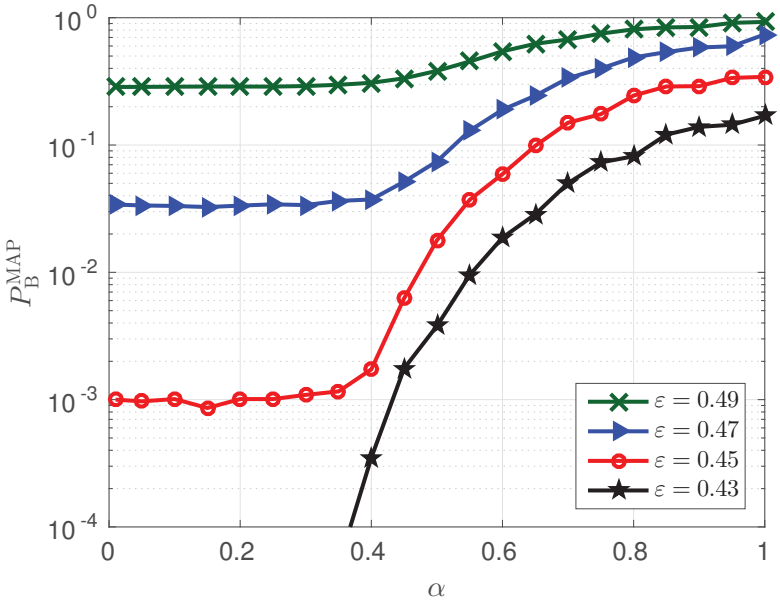


Figure 6.1 – Error probability $P_{\mathrm{B}}^{\mathrm{MAP}}$ under MAP decoding for the transmission of $\mathcal{C}_\alpha$ over the BEC(ε), when α varies from 0 to 1 with a step of 0.05 and ε is given four distinct values. The block length is $N = 2^{10}$ and the rate is $R = 0.5$. Observe that $P_{\mathrm{B}}^{\mathrm{MAP}}$ is increasing in α for all values of ε, which means that the minimum error probability is achieved by the Reed-Muller code $\mathcal{C}_\alpha\big|_{\alpha=0}$.

probability. Note that a proof of this fact would imply that Reed-Muller codes are capacity-achieving for the BEC, which is a long-standing conjecture in coding theory. We will eventually prove that Reed-Muller codes achieve capacity on erasure channels and solve this open problem in Chapter 7.

Another evidence in support of the numerical observation that $P_{\mathrm{B}}^{\mathrm{MAP}}(\alpha, \varepsilon)$ is increasing in α can be described as follows. As it has been pointed out in Section 6.1, the polar rule differs from the Reed-Muller rule in the fact that not only the number, but also the position of the 1s in $b^{(i)}$ matters in the choice of the row indices. In particular, polar codes prefer to set the 1s in the least significant bits of the binary expansion of $i - 1$. However, if one is concerned with achieving the capacity of the BEC under MAP decoding, the specific order of the 1s in the binary expansions of the row indices does not play any role. Indeed, denote by $\mathcal{F}$ the set of row indices of $F^{\otimes n}$ that are not chosen for the generator matrix of the polar code (these indices are *frozen*, since they are not used for the transmission of information bits) and let $\mathcal{F}^{\mathrm{c}}$ be its complement. Then, it is possible to arbitrarily permute the binary expansions $b^{(i)}$ ($i \in \mathcal{F}^{\mathrm{c}}$) and still get a set of row indices that yields a capacity-achieving family of codes under MAP decoding. This fact is formalized in the following proposition.

Proposition 6.2. *Denote by $\mathcal{F}^{\mathrm{c}}$ the set of row indices chosen by polar coding. Let $\pi : [n] \to [n]$ be a permutation and let P_π be the associated permutation matrix. Construct the code $\mathcal{C}_\pi$ by taking the rows of $F^{\otimes n}$ whose indices have binary expansions $P_\pi b^{(i)}$ for $i \in \mathcal{F}^{\mathrm{c}}$. Let $\varepsilon \in (0, 1)$ and denote by $P_{\mathrm{B}}^{\mathcal{D}}(\mathcal{C}_\pi)$ the error probability under the decoder $\mathcal{D}$ for the transmission of $\mathcal{C}_\pi$ over the BEC(ε). Then, $P_{\mathrm{B}}^{\mathrm{MAP}}(\mathcal{C}_\pi) \leq P_{\mathrm{B}}^{\mathrm{SC}}(\mathcal{C}_\iota)$,*

$\mathcal{C}_\iota$ being the original polar code.

Proof. As observed in [53], there exist $n!$ different representations of the polar code $\mathcal{C}_\iota$ of block length $N = 2^n$ obtained by permuting the n layers of connections. Let us apply the permutation τ to these layers and then run the SC algorithm, denoting by $P_{\mathrm{B}}^{\mathrm{SC},\tau}(\mathcal{C}_\iota)$ the error probability for the transmission over the BEC(ε). The application of the permutation τ affects the Bhattacharyya parameter $Z_n^{(i)}$ associated with the synthetic channel $W_n^{(i)}$. This Bhattacharyya parameter is now given by

$$Z_n^{(i)}(\varepsilon) = f_{\tau(b_1^{(i)})} \circ f_{\tau(b_2^{(i)})} \circ \cdots f_{\tau(b_n^{(i)})}(\varepsilon).$$

On the contrary, the generator matrix (consequently, the set $\mathcal{F}^{\mathrm{c}}$) does not change, because the code stays the same. Therefore, the probability that the SC decoder fails when applying the permutation τ to the layers of the code $\mathcal{C}_\iota$ equals the probability that the SC decoder fails when the code $\mathcal{C}_\tau$ is employed. In formulae, for any permutation τ,

$$P_{\mathrm{B}}^{\mathrm{SC},\tau}(\mathcal{C}_\iota) = P_{\mathrm{B}}^{\mathrm{SC}}(\mathcal{C}_\tau).$$

Denote by OSC the algorithm that runs SC decoding over all the $n!$ possible overcomplete representation of a polar code. When the transmission takes place over the BEC, the OSC decoder fails if and only if there exists an information bit that cannot be decoded by any of these $n!$ SC decoders. Let $P_{\mathrm{B}}^{\mathrm{OSC}}(\mathcal{C}_\pi)$ be the error probability under OSC decoding for the transmission of the code $\mathcal{C}_\pi$ over the BEC(ε). Then, $P_{\mathrm{B}}^{\mathrm{OSC}}(\mathcal{C}_\pi) \leq P_{\mathrm{B}}^{\mathrm{SC},\tau}(\mathcal{C}_\pi)$ for any τ. Taking $\tau = \pi^{-1}$ and recalling that MAP decoding minimizes the error probability, we obtain that

$$P_{\mathrm{B}}^{\mathrm{MAP}}(\mathcal{C}_\pi) \leq P_{\mathrm{B}}^{\mathrm{OSC}}(\mathcal{C}_\pi) \leq P_{\mathrm{B}}^{\mathrm{SC},\pi^{-1}}(\mathcal{C}_\pi) = P_{\mathrm{B}}^{\mathrm{SC}}(\mathcal{C}_\iota),$$

which gives us the desired result. □

In Figure 6.2, we fix the value of α and we analyze $P_{\mathrm{B}}^{\mathrm{MAP}}$ as a function of ε. It is interesting to remark that, already for $\alpha = 0.3$, the error probability for the transmission of $\mathcal{C}_\alpha$ is very close to that of random coding. Recall from Section 1.5 that random codes have a scaling exponent $\mu = 2$, whereas the scaling exponent of polar codes for the BEC is around 3.6 (see Section 2.2 for more details). Hence, the proposed interpolating family $\{\mathcal{C}_\alpha\}$ has the potential to improve the trade-off between the block length and the gap to capacity. Note also that, in principle, there is no conflict between the following two facts:

1. the error exponent of Reed-Muller codes under MAP decoding cannot be as good as that of random codes because of their minimum distance [53];

2. the scaling exponent of Reed-Muller codes can match that of random codes.

Indeed, the error exponent and the scaling exponent concern two different limits. For example, an error probability of the form $2^{-a\sqrt{N}} + 2^{-bN(C-R)^2}$ for some constants a and b yields the error exponent of polar codes and, at the same time, the scaling exponent of random codes.

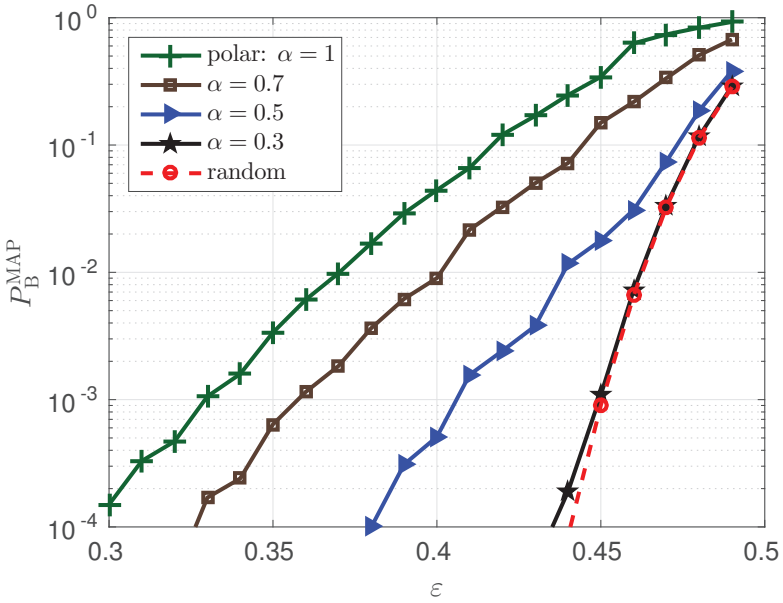


Figure 6.2 – Error probability $P_{\mathrm{B}}^{\mathrm{MAP}}$ under MAP decoding for the transmission of $\mathcal{C}_\alpha$ over the BEC(ε), when ε varies from 0.30 to 0.49 with a step of 0.005 and α is given four distinct values. The block length is $N = 2^{10}$ and the rate is $R = 0.5$. Remark that already for $\alpha = 0.3$ the error performance of $\mathcal{C}_\alpha$ is comparable to that of random codes.

6.2.2 SC Decoding

After dealing with optimal MAP decoding, let us analyze the performance of the codes $\{\mathcal{C}_\alpha\}$ under SC decoding. As can be seen in Figure 6.3 for four distinct values of ε, the error probability $P_{\mathrm{B}}^{\mathrm{SC}}(\alpha, \varepsilon)$ under SC decoding for the transmission of the code $\mathcal{C}_\alpha$ over the BEC(ε) is a decreasing function of α. Hence, the best performance is obtained using the polar code $\mathcal{C}_\alpha\big|_{\alpha=1}$. The theoretical reason for this behavior lies in the fact that $P_{\mathrm{B}}^{\mathrm{SC}}$ can be well approximated by the sum of the Bhattacharyya parameters of the synthetic channels that are selected by the polar code for the transmission of the information bits [143]. Formally, let $\mathcal{F}^{\mathrm{c}}(\alpha)$ be the set of indices selected by the polar code $\mathcal{C}_\alpha$. Then,

$$P_{\mathrm{B}}^{\mathrm{SC}}(\alpha) \leq \sum_{i \in \mathcal{F}^{\mathrm{c}}(\alpha)} Z_n^{(i)}(\varepsilon). \tag{6.2}$$

The bound (6.2) is tight when the RHS is not very large (smaller than 10^{-1} suffices) and $\sum_{i \in \mathcal{F}^{\mathrm{c}}(\alpha)} Z_n^{(i)}(\varepsilon)$ is minimized for $\alpha = 1$.

6.2.3 Something between the Extremes: List Decoding and Belief Propagation

Consider the SCL scheme introduced in [55] and denote by $P_{\mathrm{B}}^{\mathrm{SCL}}(\alpha, \varepsilon, L)$ the error probability under SCL decoding with list size L for the transmission of the polar code $\mathcal{C}_\alpha$ over the BEC(ε). Clearly, if $L = 1$, this scheme reduces to the SC algorithm

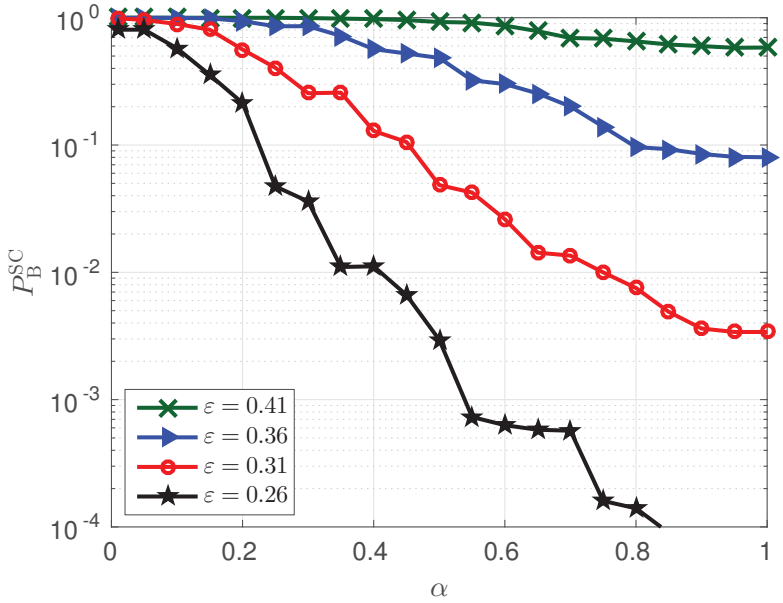
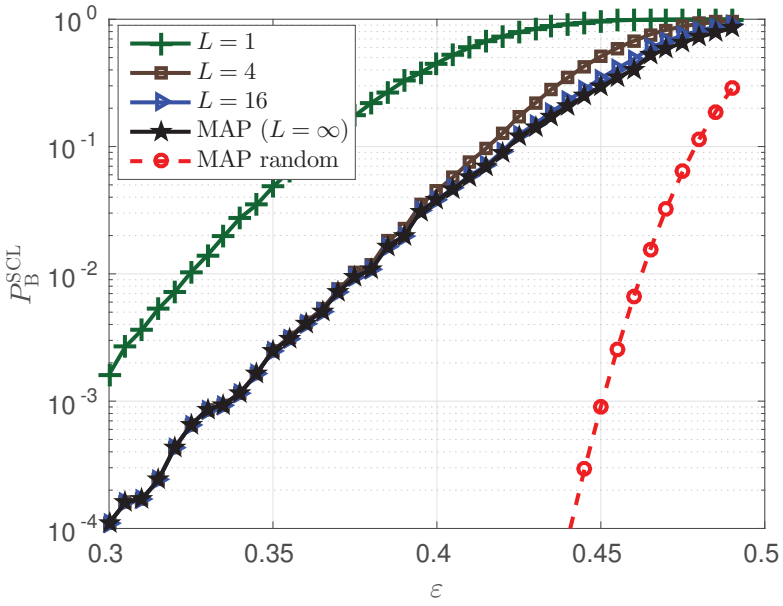


Figure 6.3 – Error probability $P_{\mathrm{B}}^{\mathrm{SC}}$ under SC decoding for the transmission of $\mathcal{C}_\alpha$ over the BEC(ε), when α varies from 0 to 1 with a step of 0.05 and ε is given four distinct values. The block length is $N = 2^{10}$ and the rate is $R = 0.5$. Observe that $P_{\mathrm{B}}^{\mathrm{SC}}$ is decreasing in α, which means that the minimum $P_{\mathrm{B}}^{\mathrm{SC}}$ is achieved by the original polar code $\mathcal{C}_\alpha\big|_{\alpha=1}$.

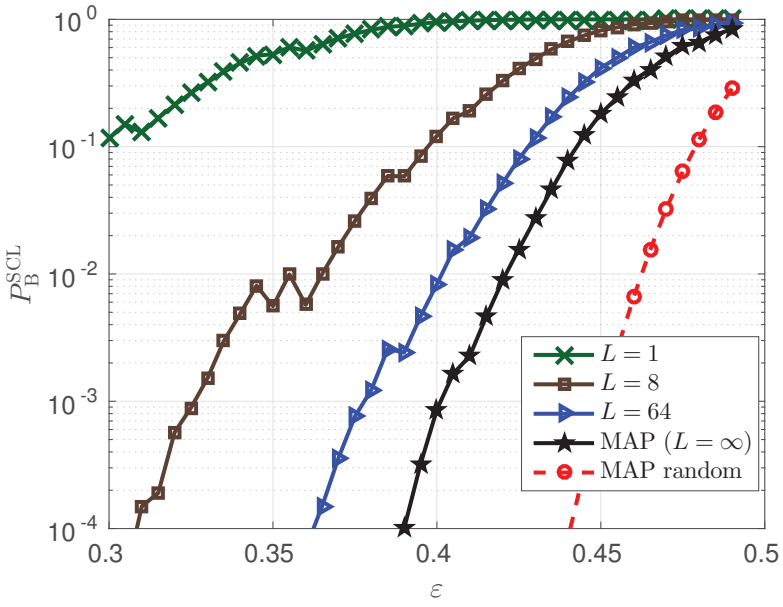
originally proposed by Arıkan, whereas for $L \geq 2^{NR}$, the SCL decoder is equivalent to the MAP decoder, as the list is big enough to contain all the possible 2^{NR} codewords. Therefore, as L increases, we gradually pass from SC decoding to MAP decoding.

If we fix α and we let L grow, $P_{\mathrm{B}}^{\mathrm{SCL}}(\alpha, \varepsilon, L)$ monotonically decreases from $P_{\mathrm{B}}^{\mathrm{SC}}(\alpha, \varepsilon)$ to $P_{\mathrm{B}}^{\mathrm{MAP}}(\alpha, \varepsilon)$. Recall that, as α goes from 1 to 0, $P_{\mathrm{B}}^{\mathrm{SC}}(\alpha, \varepsilon)$ increases and $P_{\mathrm{B}}^{\mathrm{MAP}}(\alpha, \varepsilon)$ decreases. Values of α close to 1 imply that $P_{\mathrm{B}}^{\mathrm{SCL}}(\alpha, \varepsilon, L)$ gets close to the MAP error probability for small values of the list size. If α is reduced, a bigger list size is required to obtain performance comparable to MAP decoding as the underlying SC algorithm gets worse, but $P_{\mathrm{B}}^{\mathrm{MAP}}(\alpha, \varepsilon)$ becomes significantly smaller. In other words, a smaller α implies a slower convergence (in terms of L) toward a smaller error probability. This trade-off between MAP error probability and the list size required to reach it is illustrated in Figure 6.4 for $\alpha = 0.9$ and $\alpha = 0.4$, where, for a benchmark, we represent also the error probability under MAP decoding for the transmission of random codes. Observe that if α is big (upper plot), $P_{\mathrm{B}}^{\mathrm{SCL}}$ converges to $P_{\mathrm{B}}^{\mathrm{MAP}}$ already for small values of the list size. On the contrary, if α is small (lower plot), bigger list sizes are required to get to the error probability of MAP decoding that in return becomes much smaller in value, hence much closer to the error probability of a random code. The fact that some curves are not always increasing in ε is not caused by a problem in the simulation. Indeed, the code changes with ε and, for a small variation of the channel parameter, this can lead to such unexpected effects that can be noticed also in Figures 6.5 and 6.6.

In order to show that the use of codes in $\{\mathcal{C}_\alpha\}$ significantly improves the finite-

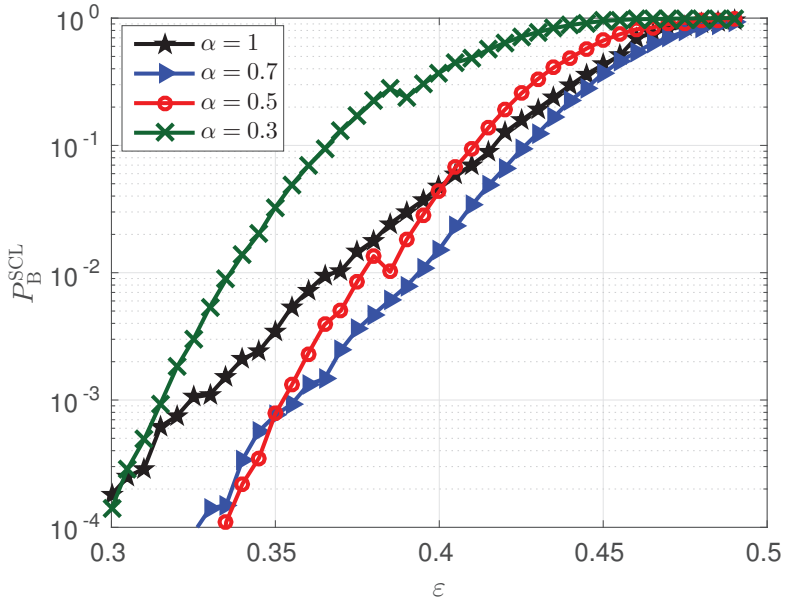


(a) $\alpha = 0.9$

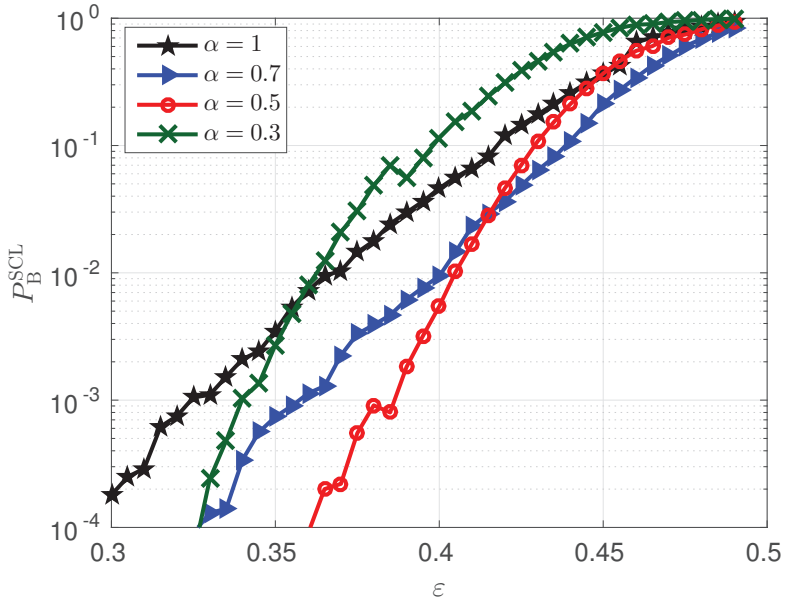


(b) $\alpha = 0.4$

Figure 6.4 – Error probability $P_{\mathrm{B}}^{\mathrm{SCL}}$ under SCL decoding for the transmission of $\mathcal{C}_\alpha$ over the BEC(ε) for different values of the list size L, when ε varies from 0.30 to 0.49 with a step of 0.005. The block length is $N = 2^{10}$ and the rate is $R = 0.5$. For a benchmark, we represent also the error probability under MAP decoding for the transmission of $\mathcal{C}_\alpha$ (in black) and for the transmission of a random code (in red).



(a) $L = 8$



(b) $L = 32$

Figure 6.5 – Error probability $P_{\mathrm{B}}^{\mathrm{SCL}}$ under SCL decoding for the transmission of $\mathcal{C}_\alpha$ over the BEC(ε), when ε varies from 0.30 to 0.49 with a step of 0.005 and for different values of α. The block length is $N = 2^{10}$ and the rate is $R = 0.5$. Already when $L = 8$ (upper plot), a performance improvement is obtained by reducing α with respect to the original polar code $\mathcal{C}_\alpha\big|_{\alpha=1}$. If the list size is increased to $L = 32$ (lower plot), the advantage in considering codes $\mathcal{C}_\alpha$ with a smaller value of the tuning parameter α is even more evident.

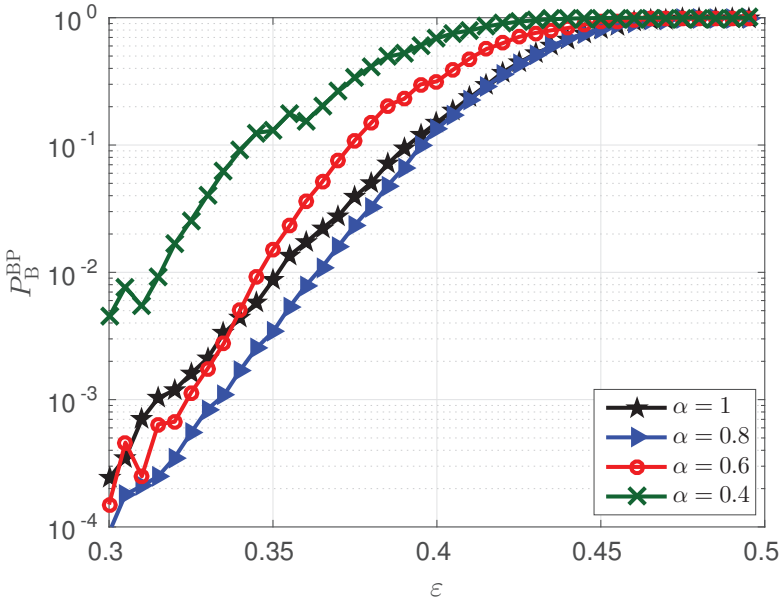


Figure 6.6 – Error probability $P_{\mathrm{B}}^{\mathrm{BP}}$ under BP decoding for the transmission of $\mathcal{C}_\alpha$ over the BEC(ε), when ε varies from 0.30 to 0.49 with a step of 0.005 and α is given four distinct values. The block length is $N = 2^{10}$ and the rate is $R = 0.5$. Remark that the optimal performance is obtained with the code $\mathcal{C}_\alpha\big|_{\alpha=0.8}$.

length performance of polar codes for practical values of the list size, fix L and consider the transmission of $\mathcal{C}_\alpha$ for different values of α. The results for $L = 8$ and $L = 32$ are represented in Figure 6.5. The code $\mathcal{C}_\alpha\big|_{\alpha=0.7}$ outperforms the original polar scheme already when $L = 8$. For $L = 32$, the improvement in performance is even more significant and, for example, the target error probability $P_{\mathrm{B}} = 10^{-3}$ can be obtained for $\varepsilon = 0.385$ if we employ $\mathcal{C}_\alpha\big|_{\alpha=0.5}$, whereas $\varepsilon = 0.325$ is required if we employ the original polar code $\mathcal{C}_\alpha\big|_{\alpha=1}$. Remark that if the target error probability to be met is very low, it is convenient to consider codes $\mathcal{C}_\alpha$ with small α, as they will be able to achieve it for higher erasure probabilities of the BEC. Indeed, observe that in the case $L = 32$, $\mathcal{C}_\alpha\big|_{\alpha=0.3}$ outperforms the original polar code for $P_{\mathrm{B}}^{\mathrm{SCL}} < 10^{-3}$. This effect is due to the fact that polar codes are not affected by error floors, as proven in Section 2.5 of this thesis.

In general, it is convenient to consider codes of the form $\mathcal{C}_\alpha$ whenever the decoding algorithm yields better results than the SC decoder. As another example, consider the case of the BP decoder. It has been already pointed out that the polar choice of the row indices to be selected from $F^{\otimes n}$ is not optimal for the BP algorithm [53, 54], but no systematic rule capable of outperforming polar codes is known. As can be seen in Figure 6.6, the interpolating family $\{\mathcal{C}_\alpha\}$ contains codes that achieve a smaller error probability than that of the original polar code $\mathcal{C}_\alpha\big|_{\alpha=1}$ for an appropriate choice of the parameter α.

6.3 Generalization to Any BMS Channel

In this section, we generalize the previous ideas to the transmission over a BMS channel W. In particular, first we propose a method for constructing the family of codes $\{\mathcal{C}_\alpha\}$, then we analyze the performance for the transmission over a binary-input AWGN channel.

6.3.1 Construction of Interpolating Family

Suppose that the transmission takes place over the BMS channel W and let $Z(W)$ be its Bhattacharyya parameter. In order to construct the interpolating family $\{\mathcal{C}_\alpha\}$, we consider the family of channels $\mathcal{W}$ ordered by degradation [44] such that the element of the family with the biggest Bhattacharyya parameter is W itself and the element of the family with the smallest Bhattacharyya parameter is the perfect channel W^{opt}, in which the output is equal to the input with probability 1. There are many ways of performing such a task. In particular, we can set

$$\mathcal{W} = \{W_\alpha : \alpha \in [0,1]\}, \tag{6.3}$$

where $W_\alpha = W$ with probability α, $W_\alpha = W^{\text{opt}}$ with probability $1-\alpha$, and the receiver knows which channel has been used. In formulae, $W_\alpha = \alpha W + (1-\alpha)W^{\text{opt}}$.

As the convex combination of BMS channels is a BMS channel, W_α is also a BMS channel with Bhattacharyya parameter $Z_\alpha = \alpha Z$. Denote by $\mathcal{C}_\alpha$ the polar code for the transmission over W_α. This is a reasonable choice for the interpolating family $\{\mathcal{C}_\alpha\}$ because of the following result that extends Proposition 6.1.

Proposition 6.3. *Let W be a BMS channel, W^{opt} be the perfect channel and $\alpha \in [0,1]$. Denote by $\mathcal{C}_\alpha$ the polar code of block length N and rate R designed for the transmission over the BMS channel $W_\alpha = \alpha W + (1-\alpha)W^{\text{opt}}$. Then, when $\alpha \to 0$, $\mathcal{C}_\alpha$ is a Reed-Muller code.*

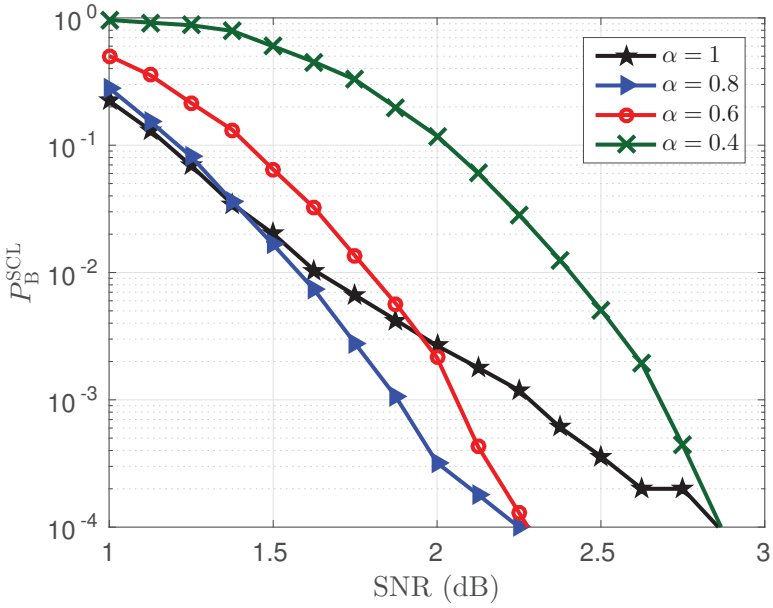
Proof. When the transmission takes place over the BMS channel W_α, the Bhattacharyya parameter $Z_n^{(i)}(W_\alpha)$ of the i-th synthetic channel $W_{\alpha,n}^{(i)}$ ($i \in [N]$) has the form (6.1), where ε is replaced by $Z_\alpha = \alpha Z$, $f_1(x) = x^2$, and $f_0(x)$ can be bounded as

$$x \le f_0(x) \le 2x - x^2, \tag{6.4}$$

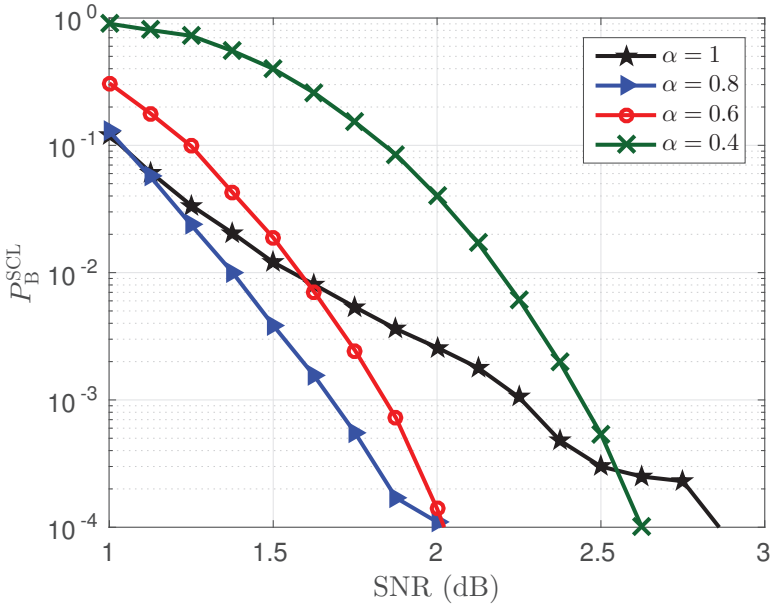
which is simply a looser version of the bound in (1.7).

Suppose that g_{j^*} is included in the generator matrix of the code, but not g_{i^*}, with $w_{\text{H}}(g_{i^*}) > w_{\text{H}}(g_{j^*})$. Then, using (6.4), $Z_n^{(i^*)}$ can be upper bounded by a polynomial in α with minimum degree $w_{\text{H}}(g_{i^*})$ and $Z_n^{(j^*)}$ can be lower bounded by a polynomial in α with minimum degree $w_{\text{H}}(g_{j^*})$. Thus, for α small enough $Z_n^{(i^*)} < Z_n^{(j^*)}$ and we reach a contradiction. □

Remark that if $W = \text{BEC}(\varepsilon)$, then $W_\alpha = \text{BEC}(\alpha\varepsilon)$. In general, there might be more natural ways to obtain the family of codes $\{\mathcal{C}_\alpha\}$, according to the particular choice of the channel W. Indeed, in Section 6.3.2 where we deal with the case of the binary-input AWGN channel, the interpolating family is constructed in a different way.



(a) $L = 8$



(b) $L = 32$

Figure 6.7 – Error probability $P_{\mathrm{B}}^{\mathrm{SCL}}$ under SCL decoding for the transmission of $\mathcal{C}_\alpha$ over the BAWGNC(σ^2), where $\sigma^2 = 0.6309$, the SNR varies from 1 to 3 with a step of 0.125 and $\alpha \in \{0.4, 0.6, 0.8, 1\}$. The block length is $N = 2^{11}$ and the rate is $R = 0.5$. The plots show the remarkable performance gain achievable by codes of the form $\mathcal{C}_\alpha$, with respect to the original polar code $\mathcal{C}_\alpha\big|_{\alpha=1}$.

In summary, the family of codes $\{\mathcal{C}_\alpha\}$ defined above is such that $\mathcal{C}_\alpha\big|_{\alpha=1}$ is the polar code designed for the transmission over the channel W and $\mathcal{C}_\alpha\big|_{\alpha=0}$ is a Reed-Muller code. Numerical simulations show that the error probability under MAP decoding is an increasing function of α. However, under SC decoding, the optimal performance is still achieved by using $\mathcal{C}_\alpha\big|_{\alpha=1}$. If we consider low-complexity decoding algorithms that get close to the error probability under MAP decoding, the finite-length performance of polar codes is significantly improved by using the code $\mathcal{C}_\alpha$ for a suitable choice of the parameter α.

6.3.2 Case Study: Binary AWGN Channel

Let W be a binary-input AWGN channel with noise variance σ^2, in short $W = \text{BAWGNC}(\sigma^2)$, and define $\mathcal{C}_\alpha$ as the polar code designed for the transmission over $W_\alpha = \text{BAWGNC}(\alpha\sigma^2)$. As $\alpha \to 0$, W_α tends to the perfect channel W^{opt} and $\mathcal{C}_\alpha$ becomes a Reed-Muller code.

In order to show the performance improvement guaranteed by the use of codes in the interpolating family $\{\mathcal{C}_\alpha\}$, consider the SCL decoder. To be coherent with the simulation setup of [55], the numerical simulations refer to codes of fixed block length $N = 2^{11}$ and rate $R = 0.5$. The number of Monte Carlo trials is $M = 10^5$. The codes are optimized for an SNR $= 2$ dB, namely, $\sigma^2 = 0.6309$ (recall that SNR $= 1/\sigma^2$). The results of Figure 6.7 are qualitatively similar to those represented in Figure 6.5 for the BEC. More specifically, for the target error probability $P_{\text{B}} = 10^{-3}$ an improvement ≥ 0.5 dB can be noticed by using the code $\mathcal{C}_\alpha\big|_{\alpha=0.8}$, rather than the original polar code $\mathcal{C}_\alpha\big|_{\alpha=1}$ when $L = 32$.

Capacity via Symmetry I: A Proven Conjecture

7

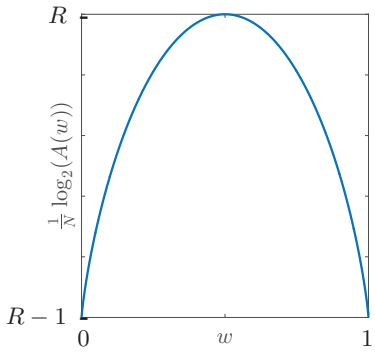
Non usare sigle commerciali & abbreviazioni etc.

Do not use acronyms & abbreviations etc.

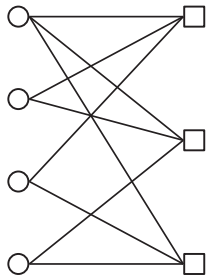
Since the landmark 1948 paper by Shannon [3], theorists have been fascinated by the challenge of constructing codes that achieve capacity, i.e., the maximum possible asymptotic rate that allows reliable communication. As discussed in Section 1.1, random coding and weight distribution, iterative coding on sparse graphs, and polarization constitute the main known techniques for proving that a family of codes is capacity-achieving (see also Figure 7.1). In this chapter, we show that also **symmetry alone guarantees asymptotically optimal performance**. A corollary of this fact is that **Reed-Muller codes are capacity-achieving** for the transmission over the BEC under MAP decoding, which settles a decade-long open problem.

The beauty of these results also lies in the simplicity of the proof, based on the interplay between three quite different ingredients: (i) *double transitivity* of the permutation group of the code, (ii) *sharp thresholds* for the measure of monotone symmetric sets, and (iii) *area theorem* for extrinsic information transfer (EXIT) functions.

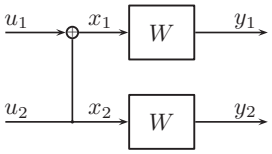
Before moving on, let me say a few words about the genesis of this work. The fact that Reed-Muller codes achieve capacity on the BEC under MAP decoding was listed as an open problem during the 2015 Simons Institute program, "Information Theory". During my visit at the Simons Institute, I had the opportunity to work on this topic with Shrinivas Kudekar, Eren Şaşoğlu, and Rüdiger Urbanke. Simultaneously, another group formed by Santhosh Kumar and Henry D. Pfister, who was also visiting the Simons Institute, independently worked on this same problem. Both the groups managed to prove the conjecture with the same technique and the two proofs were independently posted on `http://arxiv.org/` at the same time [190, 191]. As both the main result and the proof were the same, we decided to merge the submis-



(a) Random coding and weight distribution.



(b) Iterative coding on sparse graphs.



(c) Polarization.

Figure 7.1 – Pictorial representation of the three main techniques for achieving capacity: (a) the number of codewords $A(w)$ with weight w is close enough to that of an ensemble of uniformly distributed and pairwise independent codewords, i.e., $\frac{1}{N}\log_2 A(w) = R-1+h_2(w)$; (b) the threshold for reliable transmission is determined by studying the density evolution of the iterative decoding process; (c) copies of the transmission channel are polarized into synthetic channels that are either completely noisy or completely noiseless.

sions into a common joint work [192, 193]. This material also appeared in the Ph.D. thesis of Santhosh Kumar [194]. Even if the content is inevitably similar, the style of this chapter differs from [192–194]. Indeed, in order to underline the simplicity of our approach, the focus is more on the description of the proof ingredients and on how they interact to prove the conjecture, rather than on the generality of the results.

In Section 7.1, we present the main result. We introduce the three main ingredients, i.e., symmetry, sharp thresholds, and EXIT functions, in Sections 7.2, 7.3, and 7.4, respectively. Eventually, we give the proof in Section 7.5. We defer one of the proofs to the appendix in Section 7.6.

7.1 Main Result

The fundamental contribution of this chapter is given by the following theorem.

Theorem 7.1 (Doubly Transitive Codes Achieve Capacity)**.** *Let $\{\mathcal{C}_n\}$ be a sequence of codes with block lengths $N_n \to \infty$, rates $R_n \to R$, for $R \in (0,1)$, and such that the permutation group of $\mathcal{C}_n$ is doubly transitive for each n. Then, $\{\mathcal{C}_n\}$ is capacity-achieving for the transmission over the BEC under bit-MAP decoding.*

As it will be defined more rigorously in Section 7.2, the permutation group of a linear code is the set of permutations on code bits under which the code is invariant. By proving that Reed-Muller codes are doubly transitive, we are eventually able to show that they achieve capacity.

Corollary 7.1 (RM Codes Achieve Capacity)**.** *Any sequence of Reed-Muller codes with block lengths $N_n \to \infty$, and rates $R_n \to R$, for $R \in (0,1)$, is capacity-achieving for the transmission over the BEC under bit-MAP decoding.*

These results are perhaps surprising. Until the discovery of polar codes, it was unclear whether or not codes with a simple deterministic structure could be optimal in any sense [195, 196]. Furthermore, even though both polar and Reed-Muller codes derive from the Hadamard matrix $F^{\otimes n}$ defined in (1.13), the proof that polar codes achieve capacity appears unrelated to the inherent symmetry of this matrix. In contrast, the performance guarantees obtained here are a consequence only of linearity and of the structure induced by the doubly transitive permutation group.

7.2 Ingredient 1: Symmetry

Unsurprisingly, the first ingredient is symmetry. First of all, let us define the terms *doubly transitive* and *permutation group*. Denote by S_N the symmetric group on N elements and recall that, for $N \in \mathbb{N}$, $[N]$ is a shorthand for $\{1, \cdots, N\}$.

Definition 7.1 (Permutation Group)**.** *The* permutation group $\mathcal{G}$ *of a binary code* $\mathcal{C} \subseteq \{0,1\}^N$ *is defined as*

$$\mathcal{G} \triangleq \{\pi \in S_N \mid \pi(x) \in \mathcal{C} \text{ for all } x \in \mathcal{C}\}. \tag{7.1}$$

With an abuse of notation, we denote by $\pi(x)$ the vector of length N that is obtained by permuting the positions of x according to π. In words, the permutation group of a code is the set of permutations that map the code into itself.

Definition 7.2 (Transitivity)**.** *Let $\mathcal{G}$ be a permutation group. Then,*

a) $\mathcal{G}$ is transitive *if for any $i, j \in [N]$, there exists $\pi \in \mathcal{G}$ such that $\pi(i) = j$;*

b) $\mathcal{G}$ is doubly transitive *if for distinct $i, j, k \in [N]$, there exists $\pi \in \mathcal{G}$ such that $\pi(i) = i$, $\pi(j) = k$.*

The following is a classic result for Reed-Muller codes by Kasami, Lin, and Peterson [197, Corollary 4]. For the sake of completeness, its proof is in Appendix 7.6.1 and follows closely Appendix III-A of [192].

Lemma 7.1 (RM Codes Are Doubly Transitive)**.** *The permutation group $\mathcal{G}$ of the Reed-Muller code $RM(n, v)$ is doubly transitive for any $n, v \in \mathbb{N}$.*

This is the only property of Reed-Muller codes that we need in this chapter. Indeed, by using Theorem 7.1 and Lemma 7.1, the proof of Corollary 7.1 easily follows.

7.3 Ingredient 2: Sharp Thresholds

The second ingredient consists in the study of functions that experience a sharp threshold, i.e., a very quick transition from 0 to 1. On a historical note, the general method was pioneered by Margulis [198] and Russo [199]. Later, it was significantly

generalized by Talagrand in [200] and [201]. This approach has been applied to many problems in theoretical computer science with remarkable success [202–204]. In the context of coding theory, the technique was first introduced by Zémor in [205], refined further in [71], and also extended to AWGN channels in [206].

In general, threshold phenomena have been widely studied and play an important role in several fields, e.g., probability theory, statistics, physics, and computer science [207,208]. In this section, we simply state and briefly comment on a result by Tillich and Zémor in [71], as it specifically concerns coding theory, and on a general result by Friedgut and Kalai in [202], as we will use it for the proof in Section 7.5.

Consider a family of binary linear codes $\{\mathcal{C}_n\}$ and let $P_{\mathrm{B}}(N,\varepsilon)$ be the error probability for the transmission of a code of block length N over a channel with parameter ε. The parameter ε represents the quality of the transmission channel and, to be concrete, we can think of the binary erasure channel with erasure probability $\varepsilon \in [0,1]$, i.e., the BEC(ε), and to the binary symmetric channel with crossover probability $\varepsilon \in [0,1/2]$, i.e., the BSC(ε).

For any reasonable decoding algorithm, $P_{\mathrm{B}}(N,\varepsilon)$ is increasing in ε, as we expect that the error probability increases when the channel introduces more erasures or errors. Define $\varepsilon^*(N,\delta)$ as the channel parameter such that the error probability for the code of block length N is equal to δ, i.e., $P_{\mathrm{B}}(N,\varepsilon^*(N,\delta)) = \delta$. We say that the error probability experiences a sharp threshold when

$$\lim_{N\to\infty} \varepsilon^*(N,1-\delta) - \varepsilon^*(N,\delta) = 0. \tag{7.2}$$

In words, (7.2) means that the error probability passes from δ to $1-\delta$ in a window whose size vanishes with N.

The following result restates Theorem 2.3 and 5.2 of [71] in our notation.

Theorem 7.2 (Sharp Threshold for Block-MAP Decoder)**.** *Consider the transmission of a binary linear code with block length N and minimum distance $d_{\min}$ over the BEC(ε) or over the BSC(ε). Let $P_{\mathrm{B}}(N,\varepsilon)$ be the error probability under block-MAP decoding. Then,*

$$\varepsilon^*(N,1-\delta) - \varepsilon^*(N,\delta) \leq \frac{c_1(\delta)}{\sqrt{d_{\min}}}, \tag{7.3}$$

where $c_1(\delta)$ is a universal constant depending only on δ.

Note that the block-MAP decoder outputs the most likely codeword, as opposed to the bit-MAP decoder that outputs the most likely value for each bit position. We will discuss extensively how to pass from the error probability of the bit-MAP decoder to the error probability of the block-MAP decoder in Sections 8.4 and 8.5.

The result above essentially says that the error probability under block-MAP decoding experiences a sharp threshold for any family of codes such that $d_{\min}$ tends to infinity, as N grows large (see also Figure 7.2). In addition, it gives a tight upper bound on the size of the window in which the transition takes place. The upper bound is tight in the following sense. If the code sequence has a minimum distance that is linear in the block length (up to logarithmic factors), then the transition occurs in roughly $O(1/\sqrt{N})$. This is as sharp as it can be, since the random variations of the channel are already of order $1/\sqrt{N}$: the number of erasures and errors tends to a Gaussian distribution with mean $N\varepsilon$ and standard deviation $\sqrt{N\varepsilon(1-\varepsilon)}$.

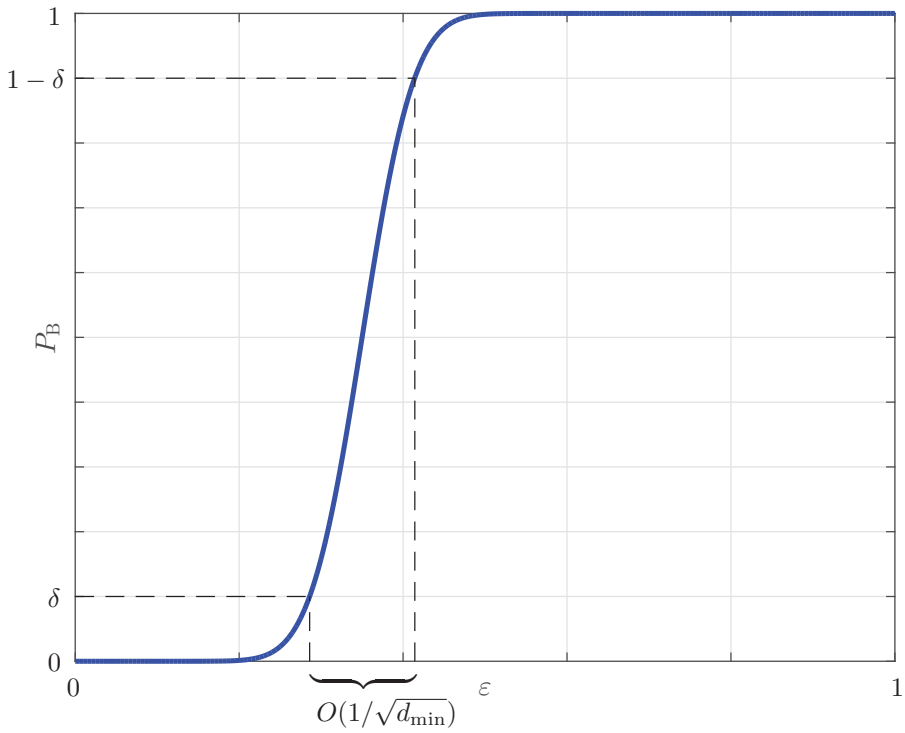


Figure 7.2 – Illustration of the result of Theorem 7.2: the error probability under block-MAP decoding P_{B} passes from δ to $1-\delta$ in a window of size $O(1/\sqrt{d_{\min}})$.

However, Theorem 7.2 does not let us to establish the location of the threshold. This happens quite frequently in theoretical computer science. On the one hand, we can apply the sharp transition framework in order to deduce that the transition width of certain functions goes to 0. On the other hand, establishing that the threshold exists, i.e., that the limit (7.2) exists, and determining its precise location is notoriously difficult [209–211].

In order to overcome these difficulties, we do not consider directly the error probability under MAP decoding, but a function closely related to it, i.e., the extrinsic information transfer (EXIT) function, as detailed in the next section. Furthermore, to show that the EXIT function exhibits a sharp transition, we resort to a more general result valid for *Bernoulli product measures* of *monotone symmetric sets*. Before stating this result, let us give some definitions. For $\omega, \omega' \in \{0,1\}^M$, we write $\omega \preceq \omega'$ when $\omega_i \leq \omega'_i$ for all $i \in [M]$.

Definition 7.3 (Monotonicity)**.** *Let $\Omega \subseteq \{0,1\}^M$. We say that Ω is* monotone *if $\omega \in \Omega$ and $\omega \preceq \omega'$ imply that $\omega' \in \Omega$.*

In words, a subset Ω of the Hamming hypercube is monotone when, by adding more 1s to one of the elements of Ω, we remain in Ω.

Definition 7.4 (Symmetry)**.** *Let $\Omega \subseteq \{0,1\}^M$. We say that Ω is* symmetric *if it transitive in the sense of Definition 7.2.*

In words, a subset Ω of the Hamming hypercube is transitive when, for any pair of indices i, $j \in [M]$, there exists a permutation that maps i into j and that keeps Ω invariant.

Definition 7.5 (Bernoulli Product Measure)**.** *Let* $\Omega \subseteq \{0,1\}^M$*. The* Bernoulli product measure *of* Ω *with parameter* ε *is denoted by* $\mu_\varepsilon(\Omega)$ *and it is defined as*

$$\mu_\varepsilon(\Omega) = \sum_{\omega \in \Omega} \varepsilon^{w_{\mathrm{H}}(\omega)} (1-\varepsilon)^{M - w_{\mathrm{H}}(\omega)}, \tag{7.4}$$

where w_{H} *denotes the Hamming weight.*

In words, the Bernoulli product measure of a subset Ω of the Hamming hypercube is the probability that a vector whose components are i.i.d. Bernoulli(ε) random variables is contained in Ω. At this point, we are ready to state Theorem 2.1 of [202] in our notation.

Theorem 7.3 (Sharp Threshold for Monotone Symmetric Sets)**.** *Let* $\Omega \subseteq \{0,1\}^M$ *be monotone and symmetric and consider the Bernoulli product measure* $\mu_\varepsilon(\Omega)$*. Define* $\varepsilon^*(\Omega, \delta)$ *as the parameter such that* $\mu_{\varepsilon^*(\Omega,\delta)}(\Omega) = \delta$*. Then,*

$$\varepsilon^*(\Omega, 1-\delta) - \varepsilon^*(\Omega, \delta) \leq c_2 \frac{\ln(1/\delta)}{\ln M}, \tag{7.5}$$

where c_2 *is a universal constant.*

In words, consider the Bernoulli product measure $\mu_\varepsilon(\Omega)$ as a function of ε. Then, if Ω is monotone and symmetric, $\mu_\varepsilon(\Omega)$ passes from δ to $1-\delta$ in a window of size $O\left(\ln(1/\delta)/\ln M\right)$.

7.4 Ingredient 3: EXIT Functions and Area Theorem

The third ingredient is the extrinsic information transfer (EXIT) function. EXIT charts were introduced by ten Brink in the context of turbo decoding as a visual tool for understanding iterative decoding [212]. For a given input bit, the EXIT function is defined to be the conditional entropy of the input bit, given the outputs associated with all other input bits. The average EXIT function is formed by averaging all of the bit EXIT functions. We note that these functions are also instrumental in the design and analysis of LDPC codes [44]. The crucial property we exploit is the so called *area theorem*, originally proven by Ashikhmin, Kramer, and ten Brink in [213] and further generalized by Méasson, Montanari, and Urbanke in [214]. This result says that the area under the average EXIT function equals the rate of the code.

For the remaining part of the chapter, we will consider the transmission over the BEC(ε). For this channel, we can define EXIT functions as follows.

Definition 7.6 (EXIT Functions)**.** *Consider the transmission of a binary linear code* $\mathcal{C}$ *of block length* N *over the BEC*(ε)*. Denote by* $X = (X_1, \cdots, X_N)$ *and* $Y = (Y_1, \cdots, Y_N)$ *the input and the output, respectively, of the channel and let* $Y_{\sim i}$ *be a shortcut for the vector* $(Y_1, \cdots, Y_{i-1}, Y_{i+1}, \cdots, Y_N)$ *that contains all outputs*

except the one at position i. Then, the EXIT function associated with bit i*, denoted by $h_i(\varepsilon)$, and the* average EXIT function*, denoted by $h(\varepsilon)$, are defined by*

$$\begin{aligned} h_i(\varepsilon) &= H(X_i \mid Y_{\sim i}), \\ h(\varepsilon) &= \frac{1}{N}\sum_{i=1}^{N} h_i(\varepsilon). \end{aligned} \tag{7.6}$$

Let us now investigate the relation between the error probability and the average EXIT function. We say that an erasure pattern covers a codeword of the code $\mathcal{C}$ when all the positions in which the codeword is 1 are erased. Let P_{b} be the error probability of the bit-MAP decoder. Clearly, for $i \in [N]$, the bit-MAP decoder recovers bit i from the output $y = (y_1, \cdots, y_N)$ if and only if the erasure pattern does not cover any codeword where bit i is non-zero. In this case, $H(X_i|Y = y) = 0$. Whenever bit i cannot be recovered uniquely, the linearity of the code implies that the set of codewords matching the unerased channel outputs has an equal number of 0s and 1s in bit position i [44, Section 3.2.2]. In this case, $H(X_i|Y = y) = 1$. Thus, the probability $P_{\mathrm{b},i}$ that the bit-MAP decoder cannot recover position i is equal to $H(X_i \mid Y)$. Since the average EXIT function can also be written in terms of entropies, it is not surprising that it is closely related to P_{b}, as proven by the following lemma (see also Lemma 3.76 of [44]).

Lemma 7.2 (EXIT Function and Bit-MAP Error Probability)**.** *Consider the transmission of a code $\mathcal{C}$ of block length N over the BEC(ε). Let $h(\varepsilon)$ be the average EXIT function and let P_{b} be the error probability under bit-MAP decoding. Then,*

$$P_{\mathrm{b}} = \varepsilon \cdot h(\varepsilon). \tag{7.7}$$

Proof. Recall that the BEC(ε) outputs a "?" with probability ε and, otherwise, leaves the input unchanged. Thus, a simple calculation shows that

$$\begin{aligned} H(X_i \mid Y) &= \mathbb{P}(Y_i = X_i) H(X_i \mid Y_{\sim i}, Y_i = X_i) + \mathbb{P}(Y_i = ?) H(X_i \mid Y_{\sim i}, Y_i = ?) \\ &= (1-\varepsilon)\cdot 0 + \varepsilon \cdot h_i(\varepsilon). \end{aligned} \tag{7.8}$$

Since $P_{\mathrm{b},i} = H(X_i \mid Y)$, by taking the average over the position i, the thesis follows. □

We consider the EXIT function rather than the error probability because there is a conservation law that says the area under the average EXIT function is always equal to the rate of the code. Therefore, the area under $h(\varepsilon)$ is an invariant for any code of a given rate: changing the code just modifies the shape of the average EXIT function, while keeping fixed the area under the curve. These considerations are formalized by the so called *area theorem*, stated and proven below (see also Lemma 3.76 and Theorem 3.82 of [44]).

Theorem 7.4 (Area Theorem)**.** *Consider the transmission of a code $\mathcal{C}$ of block length N and rate R over the BEC(ε) and let $h(\varepsilon)$ be the average EXIT function. Then,*

$$\int_0^{\varepsilon} h(x)\, dx = \frac{H(X \mid Y)}{N}, \tag{7.9}$$

where $H(X \mid Y)$ is the conditional entropy of the codeword X given the observation Y at the receiver. In particular,

$$\int_0^1 h(x)\,dx = R. \tag{7.10}$$

Proof. Although all bits are sent through the same channel BEC(ε), it is convenient to imagine that bit i is sent through a BEC(ε_i), where $\varepsilon_i = \varepsilon$ for all $i \in [N]$. Then, the derivative of the conditional entropy $H(X \mid Y)$ can be written as

$$\begin{aligned}
\frac{dH(X \mid Y(\varepsilon_1, \cdots, \varepsilon_N))}{d\varepsilon} &\overset{\text{(a)}}{=} \sum_{i=1}^{N} \frac{\partial H(X \mid Y(\varepsilon_1, \cdots, \varepsilon_N))}{\partial \varepsilon_i}\bigg|_{\varepsilon_j=\varepsilon,\, \forall j \in [N]} \\
&\overset{\text{(b)}}{=} \sum_{i=1}^{N} \frac{\partial H(X_i \mid Y(\varepsilon_1, \cdots, \varepsilon_N))}{\partial \varepsilon_i} + \frac{\partial H(X_{\sim i} \mid X_i, Y(\varepsilon_1, \cdots, \varepsilon_N))}{\partial \varepsilon_i}\bigg|_{\varepsilon_j=\varepsilon,\, \forall j \in [N]} \\
&\overset{\text{(c)}}{=} \sum_{i=1}^{N} \frac{\partial H(X_i \mid Y(\varepsilon_1, \cdots, \varepsilon_N))}{\partial \varepsilon_i}\bigg|_{\varepsilon_j=\varepsilon,\, \forall j \in [N]} \\
&\overset{\text{(d)}}{=} \sum_{i=1}^{N} \frac{\partial(\varepsilon_i \cdot H(X_i \mid Y_{\sim i}, Y_i = ?))}{\partial \varepsilon_i}\bigg|_{\varepsilon_j=\varepsilon,\, \forall j \in [N]} \\
&\overset{\text{(e)}}{=} \sum_{i=1}^{N} H(X_i \mid Y_{\sim i}, Y_i = ?) \\
&\overset{\text{(f)}}{=} Nh(\varepsilon),
\end{aligned} \tag{7.11}$$

where (a) comes from the definition of total derivative, (b) comes from the chain rule for the conditional entropy, (c) uses that $H(X_{\sim i} \mid X_i, Y)$ does not depend on ε_i, (d) follows by expanding $H(X_i \mid Y)$ as in (7.8), (e) uses that $H(X_i \mid Y_{\sim i}, Y_i = ?)$ does not depend on ε_i; and (f) follows from the definition (7.6) of average EXIT function.

By applying the fundamental theorem of calculus to (7.11), the result (7.9) immediately follows. In order to obtain (7.10), note that, when $\varepsilon = 1$, Y is independent from X, hence $H(X \mid Y) = H(X) = NR$. □

7.5 Grand Finale: The Proof

As described in the previous section, the EXIT function $h_i(\varepsilon)$ associated with bit i is the entropy of the input bit i given the outputs associated with all other input bits. This corresponds to the indirect recovery of x_i given the $N-1$ received bits $y_{\sim i}$. We denote an erasure pattern by a binary vector $\omega \in \{0,1\}^{N-1}$ that indicates the locations of the erased positions: a 1 denotes an erasure and a 0 denotes a non-erasure. The central object of our study is the set Ω_i of "bad" erasure patterns covering a codeword of $\mathcal{C}$ equal to 1 at position i. These erasure patterns are bad in the sense that they do not allow indirect recovery of the input bit i, i.e., the bit-MAP decoder cannot recover x_i from $y_{\sim i}$. Consequently, $h_i(\varepsilon)$ is encoded by

Ω_i, as it is equal to the Bernoulli product measure of this set. These concepts are formalized by the definition and the lemma that immediately follow.

Definition 7.7 (Ω_i)**.** *Consider the transmission of a binary linear code $\mathcal{C}$ of block length N over the BEC(ε). Then, Ω_i is defined as the set of all erasure patterns covering a codeword of $\mathcal{C}$ equal to* 1 *at position i, i.e.,*

$$\Omega_i = \{\omega \in \{0,1\}^{N-1} \mid x_{\sim i} \preceq \omega,\, x_i = 1 \text{ for some } x \in \mathcal{C}\}. \tag{7.12}$$

Lemma 7.3 (Ω_i Encodes $h_i(\varepsilon)$)**.** *Consider the transmission of a binary linear code $\mathcal{C}$ of block length N over the BEC(ε). Then, the set Ω_i defined in* (7.12) *contains all the erasure patterns such that it is not possible to recover the input bit x_i from the outputs $y_{\sim i}$ corresponding to all other positions. Furthermore, let $h_i(\varepsilon)$ be the EXIT function associated with bit i and $\mu_\varepsilon(\cdot)$ the Bernoulli product measure defined in* (7.4)*. Then,*

$$h_i(\varepsilon) = \mu_\varepsilon(\Omega_i). \tag{7.13}$$

Proof. As the code is linear and the channel is memoryless and symmetric, we can assume that the all-zero codeword was transmitted. Given an erasure pattern $\omega \in \{0,1\}^{N-1}$, let $\mathcal{C}'$ denote the set of all codewords x that are compatible with the observation $y_{\sim i}$, i.e., all codewords for which $x_{\sim i} \preceq \omega$.

As the code is linear, so is $\mathcal{C}'$. This implies that if there exists an $x \in \mathcal{C}'$ with $x_i = 1$, then half of all codewords in $\mathcal{C}'$ have a 0 at position i, and the other half have a 1, which means that the indirect recovery of x_i given $y_{\sim i}$ fails. Whereas, if there is no $x \in \mathcal{C}'$ with $x_i = 1$, then all compatible codewords have a 0 at position i, which means that the indirect recovery of x_i succeeds. This argument proves that Ω_i is the set of all the erasure patterns that do not allow the indirect recovery of x_i from $y_{\sim i}$.

Since the channel is memoryless, an erasure pattern ω occurs with probability $\mu_\varepsilon(\omega)$. Hence, the claim (7.13) immediately follows. □

As the discussion of Section 7.3 focuses on Bernoulli product measures of monotone symmetric sets, it is not surprising that the next step consists in proving that Ω_i is monotone and symmetric. The monotonicity follows basically by definition, whereas the symmetry comes from the fact that the code has a doubly transitive permutation group.

Lemma 7.4 (Ω_i Monotone)**.** *Consider the transmission of a binary linear code $\mathcal{C}$ of block length N over the BEC(ε). Then, the set Ω_i defined in* (7.12) *is monotone for any $i \in [N]$.*

Proof. By Definition 7.3, we need to prove that if $\omega \in \Omega_i$ and $\omega \preceq \omega'$, then $\omega' \in \Omega_i$.

If $\omega \in \Omega_i$, then there exists $x \in \mathcal{C}$ so that $x_i = 1$ and $x_{\sim i} \preceq \omega$. Since, by assumption, $\omega \preceq \omega'$, it follows that $x_{\sim i} \preceq \omega'$, which implies that $\omega' \in \Omega_i$. □

Lemma 7.5 (Ω_i Symmetric)**.** *Consider the transmission of a binary linear code $\mathcal{C}$ of block length N with* doubly transitive *permutation group over the BEC(ε). Then, the set Ω_i defined in* (7.12) *is symmetric for any $i \in [N]$.*

Proof. By Definition 7.4, we need to prove that Ω_i is transitive.

As $\mathcal{C}$ has a doubly transitive permutation group, for any $j_1, j_2 \in [N] \setminus \{i\}$, there exists a permutation $\pi : [N] \to [N]$ such that $\pi(i) = i$, $\pi(j_1) = j_2$, and $\pi(x) \in \mathcal{C}$ for any $x \in \mathcal{C}$.

Consider the permutation $\hat{\pi}$ obtained by viewing the restriction of π to $[N] \setminus \{i\}$ as a permutation from $[N-1]$ to $[N-1]$. More formally, let $S_1 : [N-1] \to [N] \setminus \{i\}$ be defined as $S_1(k) = k$ for $k \in \{1, \cdots, i-1\}$ and $S_1(k) = k+1$ for $k \in \{i, \cdots, N-1\}$. Let $S_2 : [N] \setminus \{i\} \to [N-1]$ be defined as $S_2(k) = k$ for $k \in \{1, \cdots, i-1\}$ and $S_2(k) = k-1$ for $k \in \{i+1, \cdots, N\}$. Then, $\hat{\pi}(k) = S_2(\pi(S_1(k)))$.

Note that, by changing the choice of j_1 and j_2, we generate the transitive group of permutations on $[N-1]$. Hence, in order to prove the claim, it suffices to show that if $\omega \in \Omega_i$, then $\hat{\pi}(\omega) \in \Omega_i$.

Recall that $\omega \in \Omega_i$ if there exists a codeword $x \in \mathcal{C}$ so that $x_i = 1$ and $x_{\sim i} \preceq \omega$. By construction of π, we have that $\pi(x) \in \mathcal{C}$ and $(\pi(x))_i = x_i = 1$. By construction of $\hat{\pi}$, we have that $\pi(x) \preceq \hat{\pi}(\omega)$. As a result, $\hat{\pi}(\omega) \in \Omega_i$ and the proof is complete. □

Then, we show that the EXIT functions associated with the various bits of a transitive code are all identical.

Lemma 7.6 (h_i Independent of i)**.** *Consider the transmission of a* transitive *binary linear code $\mathcal{C}$ of block length N over the BEC(ε). Let $h_i(\varepsilon)$ be the EXIT function associated with bit i. Then, $h_i(\varepsilon) = h_j(\varepsilon)$ for all $i, j \in [N]$, i.e., $h_i(\varepsilon)$ is independent of i.*

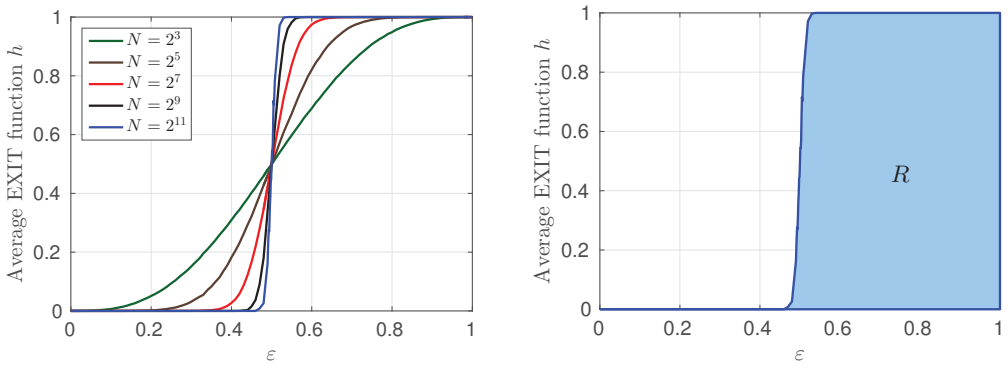
Proof. Since $\mathcal{C}$ is transitive, there exists a permutation $\pi : [N] \to [N]$ so that $\pi(i) = j$, and $\pi(x) \in \mathcal{C}$ for any $x \in \mathcal{C}$. The idea is that π maps the elements of Ω_i into the elements of Ω_j.

More specifically, pick $\omega \in \Omega_i$. Note that ω comes from an erasure pattern on the transmitted codeword, call it $\hat{\omega} \in \{0,1\}^N$, from which we have removed the observation i. Define $\hat{\pi}(\omega) \in \{0,1\}^{N-1}$ as the erasure pattern obtained by removing the observation j to $\pi(\hat{\omega})$. Since $\omega \in \Omega_i$, there exists a codeword x so that $x_i = 1$ and $x_{\sim i} \preceq \omega$. By definition of π and $\hat{\pi}$, we have that $(\pi(x))_j = 1$ and $(\pi(x))_{\sim j} \preceq \hat{\pi}(\omega)$. As a result, $\hat{\pi}(\omega) \in \Omega_j$.

As $\omega \in \Omega_i$ implies that $\hat{\pi}(\omega) \in \Omega_j$, we can think of the map $\hat{\pi}$ as going from Ω_i to Ω_j. This map is injective and it preserves the Hamming weight, i.e., $w_{\mathrm{H}}(\omega) = w_{\mathrm{H}}(\hat{\pi}(\omega))$ for any $\omega \in \Omega_i$, which implies that $\mu_\varepsilon(\Omega_i) \leq \mu_\varepsilon(\Omega_j)$. By repeating the same argument with the indices i and j exchanged, we conclude that $\mu_\varepsilon(\Omega_i) = \mu_\varepsilon(\Omega_j)$ and the thesis follows from (7.13). □

Finally, we are ready to prove our main result, i.e., Theorem 7.1. In a nutshell, the proof follows by assembling correctly the ingredients that we have described so far and, at a high level, it can be summarized by Figure 7.3. On the left side, we show that the average EXIT function becomes steeper and steeper as the block length increases, i.e., it experiences a sharp transition as predicted by Theorem 7.3. On the right side, we show that the threshold must occur at capacity from the area theorem, i.e., Theorem 7.4.

Proof of Theorem 7.1. Let $\{\mathcal{C}_n\}$ be a sequence of codes with doubly transitive permutation groups such that their block lengths $N_n \to \infty$ and rates $R_n \to R$. Consider



(a) Sharp threshold. (b) Area theorem.

Figure 7.3 – Proof by pictures of Theorem 7.1: in the left figure, we plot the average EXIT function of a sequence of codes with doubly transitive permutation group, e.g., of Reed-Muller codes, as the block length N increases; in the right figure, we indicate that the area under the average EXIT function is equal to the rate of the code.

the transmission over the BEC(ε) and bit-MAP decoding. We say that the family $\{\mathcal{C}_n\}$ achieves capacity if the error probability tends to 0 for any $R < 1 - \varepsilon$.

By Lemma 7.2, it suffices to show that the sequence of average EXIT functions tends to 0 for any $\varepsilon < 1 - R$.

By Lemma 7.4 and 7.5, the set Ω_i of Definition 7.7 is monotone and symmetric. Furthermore, by Lemma 7.3, its Bernoulli product measure $\mu_\varepsilon(\Omega_i)$ is equal to the EXIT function $h_i^{(n)}(\varepsilon)$ associated with the bit i of the code $\mathcal{C}_n$. Therefore, Theorem 7.3 bounds the window size in which $h_i^{(n)}(\varepsilon)$ passes from δ_n to $1 - \delta_n$. More specifically, we have that if $h_i^{(n)}(\overline{\varepsilon}) = 1 - \delta_n$, then $h_i^{(n)}(\underline{\varepsilon}) \leq \delta_n$ for

$$\overline{\varepsilon} = \underline{\varepsilon} + c_2 \frac{\ln(1/\delta_n)}{\ln(N_n - 1)},$$

where c_2 is a universal constant.

Lemma 7.6 implies that $h_i^{(n)}(\varepsilon)$ is independent of i, thus it is equal to the average EXIT function $h^{(n)}(\varepsilon)$ of the code $\mathcal{C}_n$. By definition of $\overline{\varepsilon}$, we have that

$$\int_0^1 h^{(n)}(\varepsilon)\, \mathrm{d}\varepsilon \geq (1 - \overline{\varepsilon})(1 - \delta_n) \geq \left(1 - \underline{\varepsilon} - c_2 \frac{\ln(1/\delta_n)}{\ln(N_n - 1)}\right)(1 - \delta_n). \tag{7.14}$$

Furthermore, Theorem 7.4 gives that

$$\int_0^1 h^{(n)}(\varepsilon)\, \mathrm{d}\varepsilon = R_n. \tag{7.15}$$

Combining (7.14) and (7.15), we obtain

$$\underline{\varepsilon} \geq 1 - \frac{R_n}{1 - \delta_n} - c_2 \frac{\ln(1/\delta_n)}{\ln(N_n - 1)}. \tag{7.16}$$

As $n \to \infty$, we have that $N_n \to \infty$ and $R_n \to R$. Thus, we can have $\delta_n \to 0$ and, at the same time, $\underline{\varepsilon}$ arbitrarily close to $1 - R$, which suffices to prove the claim. □

7.6 Appendix

7.6.1 Proof of Lemma 7.1

Proof. Take any distinct $i, j, k \in [N]$. In order to prove that $\mathcal{G}$ is doubly transitive, we will produce a $\pi \in \mathcal{G}$ such that $\pi(i) = i$ and $\pi(j) = k$.

It is well known that for any vector space with two ordered bases with m elements $(u^{(1)}, \cdots, u^{(m)})$ and $(v^{(1)}, \cdots, v^{(m)})$, there exists an invertible $m \times m$ matrix T such that $u^{(i)} = Tv^{(i)}$ for all $i \in [m]$.

Let the elements of the vector space $\{0,1\}^n$ be enumerated by $e^{(1)}, \cdots, e^{(N)}$, with $N = 2^n$, and denote by $P(n, v)$ the set of multivariate polynomial with n binary variables of degree at most v. Then, by definition of Reed-Muller code, the codewords of the code $\mathrm{RM}(n, v)$ are of the form $(f(e^{(1)}), \cdots, f(e^{(N)}))$, with $f \in P(n, v)$. Note that, as i, j, k are distinct, $e^{(j)} - e^{(i)} \neq 0_{1:n}$ and $e^{(k)} - e^{(i)} \neq 0_{1:n}$, where $0_{1:n}$ is a shorthand for the sequence of n 0s. Therefore, there exists an invertible $n \times n$ binary matrix T such that $T(e^{(j)} - e^{(i)}) = e^{(k)} - e^{(i)}$.

Let us construct $\pi \colon [N] \to [N]$ by defining $\pi(\ell)$ as the unique ℓ' such that

$$e^{(\ell')} = T(e^{(\ell)} - e^{(i)}) + e^{(i)}.$$

As T is invertible, it follows that $\pi \in S_N$, i.e., the permutation π is bijective. Also, by construction, $\pi(i) = i$ and $\pi(j) = k$.

It remains to be shown that $\pi \in \mathcal{G}$. Consider a codeword in $\mathrm{RM}(n, v)$ given by $f \in P(n, v)$. It suffices to produce a $g \in P(n, v)$ such that $g(e^{(\pi(\ell))}) = f(e^{(\ell)})$ for all $\ell \in [N]$. Let

$$g(x_1, \cdots, x_n) = f(T^{-1}[x_1, \cdots, x_n]^{\mathrm{T}} - T^{-1}e^{(i)} + e^{(i)}).$$

Then, $\mathrm{degree}(f) = \mathrm{degree}(g)$, and $g(e^{(\pi(\ell))}) = f(e^{(\ell)})$ for all $\ell \in [N]$. Therefore, $g \in P(n, v)$ and $\pi \in \mathcal{G}$. □

Capacity via Symmetry II: Generalizations

8

Non generalizzare mai.

Never generalize.

In the previous chapter, we proved that doubly transitive codes achieve capacity on the BEC under bit-MAP decoding for any rate in $(0, 1)$. In this chapter, we focus on several extensions of this theorem.

In Section 8.1, we briefly review most of the generalizations described in [192]. In Section 8.2, we outline the results presented in this chapter. In particular, in Section 8.3, we analyze Reed-Muller codes in the low-rate and high-rate regimes. The discussion on rates converging to 0 is in Section VI-E of [192], whereas the part on rates converging to 1 is new and unpublished. Then, we study how to strenghten results regarding the bit-MAP decoder to the block-MAP decoder. For the sake of completeness, in Section 8.4 we revise the proof by Kumar and Pfister that Reed-Muller codes achieve capacity under block-MAP decoding [192]. In Section 8.5, we present the main technical contribution of this chapter[1]: we describe a general framework for passing from the bit-MAP error probability to the block-MAP error probability. We defer the proofs of two intermediate results to the Appendix in Section 8.6.

8.1 Related Work

In addition to Reed-Muller codes, several families of codes of great interest are doubly transitive, hence capacity-achieving. Specifically, affine-invariant codes, BCH codes, and quadratic-residue codes are analyzed in detail in Section V of [192]. Furthermore, as pointed out in Section VI-D of [192], it is straightforward to extend the results for binary linear codes to linear codes on $\mathbb{F}_q$ transmitted over the q-ary erasure channel.

[1] This material is based on joint work with S. Kudekar, S. Kumar, H. D. Pfister, and R. Urbanke [215].

Finally, as stated in Section VI-C of [192], it is possible to link the performance of Reed-Muller codes over the binary erasure channel to their performance over the binary symmetric channel. More specifically, Theorem 8 of [68] proves that if the code $\mathrm{RM}(n, n-(t+1))$ can correct a certain erasure pattern, then the code $\mathrm{RM}(n, n-(2t+2))$ can correct an error pattern with the same support. These error patterns can even be corrected efficiently: by Corollary 14 of [216], there exists a deterministic algorithm that runs in time at most n^4 and is able to correct $(1/2 - o(1))2^n$ random errors in the code $\mathrm{RM}(n, o(\sqrt{n}))$ with probability $1 - o(1)$.

8.2 Main Results

Our contributions in this chapter can be summarized as follows.

Low-rate and high-rate regimes. A key result of [68, 69] is that Reed-Muller codes can correct almost all erasure patterns up to the capacity limit for rates approaching either 0 or 1 with sufficient speed. With the proof technique introduced in Chapter 7, we can prove that Reed-Muller codes are capacity-achieving in strictly different regimes, i.e., for different speeds of convergence.

From bit-MAP to block-MAP. We describe in detail how to strengthen results regarding the bit-MAP error probability to the block-MAP error probability. First, we present the proof by Kumar and Pfister that Reed-Muller codes achieve capacity over the BEC under block-MAP decoding [192]: by using additional symmetries of the code, one can prove a stronger result on the transition speed of the bit-MAP error probability, hence a simple union bound enables us to conclude that also the block-MAP error probability vanishes. This approach relies on the fact that the transmission occurs over an erasure channel.

One of the main open problems consists in proving that Reed-Muller codes (and, in general, codes with sufficient symmetry) are capacity-achieving for the transmission over any BMS channel. We present some progress towards such a goal by describing a general framework to pass from the bit-MAP error probability to the block-MAP error probability. The crucial idea consists in showing that, even if Reed-Muller codes have a minimum distance that scales as the square root of the block length, the codewords that actually yield errors under MAP decoding have almost linear weight. To prove such a fact, we need to provide a careful bound on the weight distribution.

8.3 Low-Rate and High-Rate Regimes

8.3.1 Rates Converging to 0

Let $\{\mathrm{RM}(n, v_n)\}$ be a sequence of Reed-Muller codes with block lengths $N_n = 2^n$ and rates $R_n \to 0$ sufficiently fast. Assume that the code $\mathrm{RM}(n, v_n)$ is transmitted over a $\mathrm{BEC}(\varepsilon_n)$ and let $P_{\mathrm{B}}^{(n)}$ be the error probability under block-MAP decoding. Then, we say that the family of codes achieves capacity if, for any $\zeta > 0$,

$$P_{\mathrm{B}}^{(n)} \to 0, \quad \text{for any } 0 \leq \varepsilon_n < 1 - (1+\zeta)R_n. \tag{8.1}$$

In Corollary 44 of [68], it is proved that (8.1) holds for $v_n \leq \eta n$ with $\eta = O(1/\ln(1/\zeta))$. Therefore, a necessary condition for this result is that $R_n = O(N_n^{-\kappa})$ for some arbitrarily small and fixed $\kappa > 0$. With the proof technique developed in the previous chapter, we can show that Reed-Muller codes achieve capacity for a different decay rate of R_n.

Theorem 8.1 (Rates $\to 0$)**.** *Let $\{RM(n, v_n)\}$ be a sequence of Reed-Muller codes with block lengths $N_n = 2^n \to \infty$ and rates $R_n \to 0$ and consider the transmission over the family of channels $\{BEC(\varepsilon_n)\}$. Assume that*

$$R_n \ln N_n \to \infty. \tag{8.2}$$

Then, $\{RM(n, v_n)\}$ achieves capacity under bit-MAP decoding.

Proof. Let $P_{\mathrm{b}}^{(n)}$ be the error probability under bit-MAP decoding for the transmission of the code $\mathrm{RM}(n, v_n)$ over the $\mathrm{BEC}(\varepsilon_n)$. From the argument in the proof of Theorem 7.1, we have that $P_{\mathrm{b}}^{(n)} \leq \varepsilon_n \delta_n$ for

$$\varepsilon_n = 1 - \frac{R_n}{1-\delta_n} - c_2 \frac{\ln(1/\delta_n)}{\ln(N_n - 1)}, \tag{8.3}$$

where c_2 is a universal constant.

We can now rewrite (8.3) as

$$\varepsilon_n = 1 - (1 + \zeta_n) R_n,$$

with

$$\zeta_n = \frac{\delta_n}{1-\delta_n} + c_2 \frac{\ln(1/\delta_n)}{R_n \cdot \ln(N_n - 1)}.$$

Set $\delta_n = 1/\ln(R_n \cdot \ln(N_n - 1))$. Then, by applying the hypothesis (8.2), we obtain that $P_{\mathrm{b}}^{(n)} \to 0$ and $\zeta_n \to 0$, which implies the desired result. □

Remark 8.1 (Comparison between Theorem 8.1 and Corollary 44 of [68])**.** *Note that if $R_n = O(N_n^{-\kappa})$, then (8.2) does not hold. Hence, the regimes considered by Corollary 44 of [68] and by Theorem 8.1 of this thesis do not overlap. In particular, we consider a case in which R_n converges to 0 much slower than in [68].*

To make more clear the comparison between these two regimes, let us rewrite the condition (8.2) in terms of the parameters v_n and n of the Reed-Muller code. As $R_n \to 0$, we have that $v_n \leq n/2$. Thus, by Lemma 4.7.2 of [217], we can bound the rate R_n by

$$\frac{2^{n h_2(\frac{v_n}{n})}}{2^n \sqrt{2n}} \leq R_n = \frac{\sum_{i=0}^{v_n} \binom{n}{i}}{2^n} \leq \frac{2^{n h_2(\frac{v_n}{n})}}{2^n}, \tag{8.4}$$

where $h_2(x) = -x \log_2 x - (1-x)\log_2(1-x)$. Furthermore, recall that the Taylor expansion of the function $h_2(\cdot)$ around $1/2$ is given by

$$h_2\left(\frac{1}{2} - x\right) = 1 - \frac{2}{\ln 2} x^2 + o(x^2).$$

Hence, after some manipulations, we finally obtain that the condition (8.2) *is equivalent to*

$$v_n = \frac{n}{2} - o(\sqrt{n \log_2 n}). \tag{8.5}$$

Recall also that the regime in which $R_n \to R$*, for any* $R \in (0,1)$*, corresponds to* $v_n = n/2 + \Theta(\sqrt{n})$*.*

8.3.2 Rates Converging to 1

Let $\{\mathrm{RM}(n, v_n)\}$ be a sequence of Reed-Muller codes with block lengths $N_n = 2^n$ and rates $R_n \to 1$ sufficiently fast. Again, assume that the code $\mathrm{RM}(n, v_n)$ is transmitted over a $\mathrm{BEC}(\varepsilon_n)$ and let $P_{\mathrm{B}}^{(n)}$ be the error probability under block-MAP decoding. Then, we say that the family of codes achieves capacity if, for any $\zeta > 0$,

$$P_{\mathrm{B}}^{(n)} \to 0, \quad \text{for any } 0 \leq \varepsilon_n < (1 - R_n)(1 - \zeta). \tag{8.6}$$

In Corollary 45 of [68], it is proved that Reed-Muller codes are capacity-achieving for $n - v_n = O(\sqrt{n/\log_2 n})$. With the proof technique developed in the previous chapter, we can show that Reed-Muller codes achieve capacity for a different decay rate of R_n.

Theorem 8.2 (Rates $\to 1$)**.** *Let* $\{RM(n, v_n)\}$ *be a sequence of Reed-Muller codes with block lengths* $N_n = 2^n \to \infty$ *and rates* $R_n \to 1$ *and consider the transmission over the family of channels* $\{BEC(\varepsilon_n)\}$*. Fix* $\bar{R}_n = 1 - R_n$ *and suppose that*

$$\frac{\bar{R}_n \ln N_n}{\ln(1/\bar{R}_n)} \to \infty. \tag{8.7}$$

Then, $\{RM(n, v_n)\}$ *achieves capacity under bit-MAP decoding.*

Proof. From the argument in the proof of Theorem 7.1, we deduce that the bit-MAP error probability $P_{\mathrm{b}}^{(n)}$ is upper bounded by $\varepsilon_n \delta_n$, where ε_n is given by (8.3).

We can now rewrite (8.3) as

$$\varepsilon_n = (1 - R_n)(1 - \zeta_n),$$

with

$$\begin{aligned} \zeta_n &= \frac{R_n}{1 - R_n} \cdot \frac{\delta_n}{1 - \delta_n} + c_2 \frac{\ln(1/\delta_n)}{(1 - R_n) \cdot \ln(N_n - 1)} \\ &= \frac{\delta_n}{\bar{R}_n} \cdot \frac{R_n}{1 - \delta_n} + c_2 \frac{\ln(1/\delta_n)}{\bar{R}_n \cdot \ln(N_n - 1)}, \end{aligned}$$

where $\bar{R}_n = 1 - R_n$. Set $\delta_n = \bar{R}_n / \ln(\bar{R}_n \cdot \ln(N_n - 1))$. Then, by applying the hypothesis (8.7), we can verify that $P_{\mathrm{b}}^{(n)} \to 0$ and $\zeta_n \to 0$, which implies the desired result. □

Remark 8.2 (Comparison between Theorem 8.2 and Corollary 45 of [68])**.** *Let us apply* (8.4) *to* $\bar{R}_n$*, rather than to* R_n*. Then, after some further calculations, we verify that if* $n - v_n = O(\sqrt{n/\log_2 n})$*, then the condition* (8.7) *does not hold. Hence, the regimes considered by Corollary 45 of [68] and by Theorem 8.2 of this thesis do*

not overlap. In particular, we consider a case in which R_n converges to 1 much slower than in [68].

To make more clear the comparison between these two regimes, let us rewrite the condition (8.7) *in terms of the parameters v_n and n of the Reed-Muller code. By using a procedure similar to that of Remark 8.1, we conclude that* (8.7) *is equivalent to*

$$v_n = \frac{n}{2} + o(\sqrt{n \log_2 n}). \tag{8.8}$$

8.4 From Bit-MAP to Block-MAP via Sharper Thresholds for BEC

First of all, let us recall the difference between bit-MAP and block-MAP decoding.

- The *bit-MAP decoder* outputs the most likely bit value for each position, and its bit error probability is denoted by P_{b}.
- The *block-MAP decoder* outputs the most likely codeword, and its block error probability is denoted by P_{B}.

Note that P_{b} is the *bit error probability* of the bit-MAP decoder and P_{B} is the *block error probability* of the block-MAP decoder. To avoid being verbose, we will simply refer to P_{b} and P_{B} as the error probabilities of the bit-MAP and block-MAP decoders, respectively, without specifying whether it is a bit error probability or a block error probability.

The main result of this section consists in showing that Reed-Muller codes achieve capacity over the BEC also under block-MAP decoding and it is stated as follows.

Theorem 8.3 (From Bit-MAP to Block-MAP for BEC)**.** *Any sequence of Reed-Muller codes with block lengths $N_n \to \infty$, and rates $R_n \to R$, for $R \in (0,1)$, is capacity-achieving for the transmission over the BEC under block-MAP decoding.*

The idea of the proof consists in providing a stronger bound on the window size in which the average EXIT function transitions from δ to $1-\delta$. Recall that Theorem 7.3 yields a window size that is $O\left(\ln(1/\delta)/\ln N\right)$. This implies that P_{b} can decay as $N^{-\gamma}$ for a fixed and arbitrarily small $\gamma > 0$. In order to speed up the decay of P_{b} to, for example, N^{-2}, it would suffice to show that the size of the transition window is $O\left(\ln(1/\delta)/(w_N \cdot \ln N)\right)$, where $w_N \to \infty$ as $N \to \infty$.

To prove such a fact, we resort to the framework developed by Bourgain and Kalai in [218], where additional properties of the permutation group are exploited in order to provide stronger results on threshold intervals. Analogously to Definition 7.1, the permutation group of a set Ω is defined as the group of permutations that leave Ω invariant. The result that we need is stated below and it is proved in Appendix 8.6.1.

Theorem 8.4 (Sharper Threshold for More Symmetric Sets)**.** *Let $\Omega \subseteq \{0,1\}^M$ be a monotone symmetric set. In addition, assume that the permutation group of Ω is isomorphic to the general linear group $GL(m, \mathbb{F}_2)$ of degree m in $\mathbb{F}_2$, where $M = 2^m$.*

Consider the Bernoulli product measure $\mu_\varepsilon(\Omega)$ and define $\varepsilon^*(\Omega,\delta)$ as the parameter such that $\mu_{\varepsilon^*(\Omega,\delta)}(\Omega)=\delta$. Then, there exists a universal constant C_1 such that

$$\varepsilon^*(\Omega,1-\delta)-\varepsilon^*(\Omega,\delta)\leq C_1\frac{\ln(1/\delta)}{\ln(\ln M)\cdot\ln M}, \tag{8.9}$$

provided that $\varepsilon^*(\Omega,\delta)(1-\varepsilon^*(\Omega,1-\delta))$ is bounded away from 0.

Luckily enough, Reed-Muller codes have the additional symmetries required by Theorem 8.4.

Lemma 8.1 (More Symmetries for RM Codes)**.** *For the code RM(n,v), consider the set Ω_N of Definition 7.7 and let $\mathcal{G}_N$ be its permutation group. Then, $\mathcal{G}_N$ contains a transitive subgroup isomorphic to $GL(n,\mathbb{F}_2)$.*

The proof of the lemma above follows closely Appendix III-B of [192] and it is reproduced with our notation in Appendix 8.6.2. Eventually, we are ready to show the main result of this section.

Proof of Theorem 8.3. Let $\{\mathrm{RM}(n,v_n)\}$ be a sequence of Reed-Muller codes with block lengths $N_n=2^n\to\infty$ and rates $R_n\to R$, for $R\in(0,1)$. Consider the transmission over the BEC(ε) and block-MAP decoding. We say that the family $\{\mathrm{RM}(n,v_n)\}$ achieves capacity if the error probability $P_{\mathrm{B}}^{(n)}$ tends to 0 for any $R<1-\varepsilon$.

The proof is similar to the one of Theorem 7.1 at the end of Section 7.5. The set Ω_N is monotone and symmetric and its Bernoulli product measure is equal to the EXIT function $h_N^{(n)}(\varepsilon)$ associated with the last bit. Furthermore, as Reed-Muller codes are transitive, $h_N^{(n)}(\varepsilon)$ is equal to the average EXIT function $h^{(n)}(\varepsilon)$.

By Lemma 8.1, the permutation group of Ω_N contains a transitive subgroup isomorphic to $\mathrm{GL}(n,\mathbb{F}_2)$. Therefore, by Theorem 8.4, we have that if $h_N^{(n)}(\overline{\varepsilon})=1-\delta_n$, then $h_N^{(n)}(\underline{\varepsilon})\leq\delta_n$ for

$$\overline{\varepsilon}=\underline{\varepsilon}+C_1\frac{\ln(1/\delta_n)}{\ln(\ln(N_n-1))\cdot\ln(N_n-1)},$$

provided that $\underline{\varepsilon}(1-\overline{\varepsilon})$ is bounded away from 0.

By using the same argument based on the area theorem, i.e., Theorem 7.4, we conclude that

$$\underline{\varepsilon}\geq 1-R_n-\delta_n-C_1\frac{\ln(1/\delta_n)}{\ln(\ln(N_n-1))\cdot\ln(N_n-1)}. \tag{8.10}$$

As $n\to\infty$, we have that $N_n\to\infty$ and $R_n\to R$. Thus, we can take $\delta_n=1/N_n^2$ and, at the same time, $\underline{\varepsilon}$ arbitrarily close to $1-R$. As a sanity check, note that $\underline{\varepsilon}(1-\overline{\varepsilon})$ is bounded away from 0, since $R\in(0,1)$.

As a result, the thesis follows from the chain of inequalities below:

$$\begin{aligned}
P_{\mathrm{B}}^{(n)} &\overset{\text{(a)}}{\leq} N_n P_{\mathrm{b}}^{(n)} \\
&\overset{\text{(b)}}{=} N_n\cdot\varepsilon\cdot h^{(n)}(\varepsilon) \\
&\leq N_n\cdot\varepsilon\cdot\delta_n \\
&\overset{\text{(c)}}{\leq} \frac{\varepsilon}{N_n}\to 0,
\end{aligned}$$

where the inequality (a) comes from the fact that, for the transmission over the BEC, if the bit-MAP decoder cannot decode at least one of the bits, then the block-MAP decoder cannot decode; the equality (b) comes from Lemma 7.2; and the inequality (c) by our choice of δ_n. □

8.5 From Bit-MAP to Block-MAP via Weight Distribution for BMS Channels

8.5.1 Statement and Proof of Main Theorem

Theorem 8.5 (From Bit-MAP to Block-MAP for BMS Channels)**.** *Consider a sequence of Reed-Muller codes with block lengths $N_n \to \infty$ and rates $R_n \to R$, for $R \in (0,1)$. Assume that each code is transmitted over a BMS channel with Bhattacharyya parameter $z \in (0,1)$ and that the error probability of the bit-MAP decoder is $O(N_n^{-\gamma})$, for a fixed $\gamma > 0$. Then, the error probability of the block-MAP decoder tends to* 0*.*

Consider the special case of transmission over the BEC. As pointed out in the previous section, Theorem 7.1 implies that P_{b} is $O(N^{-\gamma})$, for a fixed $\gamma > 0$. Thus, the result above immediately implies that Reed-Muller codes achieve capacity under block-MAP decoding without resorting to the framework of [218].

In order to prove Theorem 8.5, it is useful to introduce a randomized version of the bit-MAP and block-MAP decoders.

- The *randomized bit-MAP decoder* outputs each bit value according to its posterior probability, and its bit error probability is denoted by $P_{\mathrm{b,r}}$.
- The *randomized block-MAP decoder* outputs each codeword according to its posterior probability, and its block error probability is denoted by $P_{\mathrm{B,r}}$.

The error probabilities of the MAP decoders are related to the error probabilities of their randomized counterparts by the following lemma that is proved in Section 8.5.2.

Lemma 8.2 (MAP vs. Randomized MAP)**.** *Consider the transmission of a code $\mathcal{C}$ over a BMS channel and let P_{b}, P_{B}, $P_{\mathrm{b,r}}$ and $P_{\mathrm{B,r}}$ be the error probabilities of the bit-MAP, block-MAP, randomized bit-MAP and randomized block-MAP decoders. Then, the following inequalities hold:*

$$P_{\mathrm{b}} \leq P_{\mathrm{b,r}} \leq 2 \cdot P_{\mathrm{b}}, \tag{8.11}$$

$$P_{\mathrm{B}} \leq P_{\mathrm{B,r}} \leq 2 \cdot P_{\mathrm{B}}. \tag{8.12}$$

A crucial point in the proof of the main result of this section is that for any Reed-Muller code of sufficiently large block length and any $\beta > 0$, the codewords at distance at most $N^{1-\beta}$ from the transmitted codeword have a negligible effect on the block error probability under randomized block-MAP decoding. This means that, even if Reed-Muller codes have only a minimum distance of $\Theta(\sqrt{N})$, the codewords that really produce errors under MAP decoding have almost linear weight. These concepts are formalized by the following lemma.

Lemma 8.3 (Small Distances Do Not Count)**.** *Consider a sequence of Reed-Muller codes with block lengths $N_n \to \infty$ and rates $R_n \to R$, for $R \in (0,1)$. Assume that each code is transmitted over a BMS channel with Bhattacharyya parameter $z \in (0,1)$ and fix any $\beta > 0$. Then, the probability that the randomized block-MAP decoder outputs an incorrect codeword at Hamming distance at most $N_n^{1-\beta}$ from the transmitted codeword tends to* 0*.*

The proof of Lemma 8.3 is deferred to Section 8.5.2 and relies on an upper bound on the weight distribution of Reed-Muller codes. Recall that the codewords of the code $\mathrm{RM}(n,v)$ are given by the evaluations of the polynomials in n variables of degree at most v over $\mathbb{F}_2$. With an abuse of notation, we can think of $\mathrm{RM}(n,v)$ as the collection of such polynomials $f : \mathbb{F}_2^n \to \mathbb{F}_2$. The normalized weight of a function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is the normalized number of 1s in it, i.e.,

$$\mathrm{wt}(f) = \frac{1}{2^n} |\{x \in \mathbb{F}_2^n : f(x) = 1\}|.$$

The cumulative weight distribution of $\mathrm{RM}(n,v)$ at a normalized weight $\alpha \in [0,1]$ is denoted by $W_{n,v}(\alpha)$ and is defined as the number of codewords whose normalized weight is at most α, i.e.,

$$W_{n,v}(\alpha) = |\{f \in \mathrm{RM}(n,v) : \mathrm{wt}(f) \le \alpha\}|. \tag{8.13}$$

The study of the weight distribution of Reed-Muller codes is a classic problem in coding theory [219–221] and it culminated in the asymptotically tight bounds for fixed order v and asymptotic n by Kaufman, Lovett, and Porat [222]. These bounds were further improved in [68, 69]. More specifically, we need an additional refinement of Theorem 3.1 of [222], stated below and proven in Section 8.5.2.

Lemma 8.4 (Upper Bound on Weight Distribution)**.** *Consider the code $RM(n,v)$. Pick an integer $\ell \in [v-1]$ and $\epsilon \in (0, 1/2]$. Set $\alpha = 2^{-\ell}(1-\epsilon)$. Then,*

$$W_{n,v}(\alpha) \le (1/\epsilon)^{C_2 (v+2)^2 (n\ell + \sum_{i=0}^{v-\ell} \binom{n-\ell}{i})},$$

where C_2 is a universal constant.

Finally, we can proceed with the proof of Theorem 8.5.

Proof of Theorem 8.5. The quantities N, P_{b}, P_{B}, $P_{\mathrm{b,r}}$, and $P_{\mathrm{B,r}}$ that appear in this proof are all indexed by n, but we drop the index to avoid cluttering. Furthermore, let $P_{\mathrm{B,r}}^{\mathrm{l}}$ be the probability that the randomized block-MAP decoder outputs an incorrect codeword whose Hamming distance from the transmitted codeword is *at most* $N^{1-\gamma/2}$. Similarly, let $P_{\mathrm{B,r}}^{\mathrm{h}}$ be the probability that the randomized block-MAP decoder outputs an incorrect codeword whose Hamming distance from the transmitted codeword is *at least* $N^{1-\gamma/2}$. Then,

$$P_{\mathrm{B}} \le P_{\mathrm{B,r}} = P_{\mathrm{B,r}}^{\mathrm{l}} + P_{\mathrm{B,r}}^{\mathrm{h}},$$

where the inequality comes from (8.12). By Lemma 8.3, we have that $P_{\mathrm{B,r}}^{\mathrm{l}}$ tends to 0. Hence, in order to prove the claim, it suffices to show that also $P_{\mathrm{B,r}}^{\mathrm{h}}$ tends to 0.

To do so, we first upper bound $P_{\mathrm{B,r}}^{\mathrm{h}}$ as a function of $P_{\mathrm{b,r}}$, by adapting the proof of (13.51) in [223]. Let $x = (x_1, \cdots, x_N)$ denote a codeword and y the channel output. By definition, the randomized bit-MAP decoder outputs the value in position i according to the distribution $p(x_i \mid y)$. However, we can also draw a sample from $p(x_i \mid y)$ by first sampling from the joint distribution $p(x \mid y)$ and then by discarding all positions except position i. Now, let A be the event in which the Hamming distance between x and the transmitted codeword is at least $N^{1-\gamma/2}$. Thus,

$$\begin{aligned} P_{\mathrm{b,r}} &= \mathbb{P}(A) \cdot \mathbb{P}(\text{bit error} \mid A) + \mathbb{P}(A^{\mathrm{c}}) \cdot \mathbb{P}(\text{bit error} \mid A^{\mathrm{c}}) \\ &\geq \mathbb{P}(A) \cdot \mathbb{P}(\text{bit error} \mid A) \geq P_{\mathrm{B,r}}^{\mathrm{h}} \cdot N^{-\gamma/2}, \end{aligned} \tag{8.14}$$

where A^{c} denotes the complement of A. To prove the last inequality, note that $\mathbb{P}(A) = P_{\mathrm{B,r}}^{\mathrm{h}}$ and that, since x has Hamming distance at least $N^{1-\gamma/2}$ from the transmitted codeword, at least a fraction $N^{-\gamma/2}$ of the bits in x is decoded incorrectly by the randomized bit-MAP decoder. Finally,

$$\begin{aligned} P_{\mathrm{B,r}}^{\mathrm{h}} &\overset{\text{(a)}}{\leq} P_{\mathrm{b,r}} \cdot N^{\gamma/2} \\ &\overset{\text{(b)}}{\leq} 2 \cdot P_{\mathrm{b}} \cdot N^{\gamma/2}, \end{aligned}$$

where the inequality (a) is obtained from (8.14), and the inequality (b) from (8.11). Since, by hypothesis, P_{b} is $O(N^{-\gamma})$, the result is readily proved. □

8.5.2 Proof of Auxiliary Lemmas and Further Remarks

We start by proving Lemma 8.2.

Proof of Lemma 8.2. The inequalities $P_{\mathrm{b}} \leq P_{\mathrm{b,r}}$ and $P_{\mathrm{B}} \leq P_{\mathrm{B,r}}$ follow from the fact that the MAP decoder is, by definition, an optimal decoder in the sense that it minimizes the error probability.

In order to prove the other inequality in (8.12), let $x \in \mathcal{C}$ denote a codeword, $y \in \mathcal{Y}$ the channel output, and $\hat{x}_{\mathrm{B}}(y)$ the estimate provided by the block-MAP decoder given the channel output y. Then, we can rewrite P_{B} as

$$\begin{aligned} P_{\mathrm{B}} &= \sum_{x \in \mathcal{C}} p(x) \sum_{y \in \mathcal{Y}} p(y \mid x) \mathbb{P}(\hat{x}_{\mathrm{B}}(y) \neq x) \\ &= \sum_{y \in \mathcal{Y}} p(y) \sum_{x \in \mathcal{C}} p(x \mid y) \mathbb{P}(\hat{x}_{\mathrm{B}}(y) \neq x) \\ &\overset{\text{(a)}}{=} \sum_{y \in \mathcal{Y}} p(y) \sum_{x \in \mathcal{C} \setminus \hat{x}_{\mathrm{B}}(y)} p(x \mid y) \\ &= \sum_{y \in \mathcal{Y}} p(y) \left(1 - p(\hat{x}_{\mathrm{B}}(y) \mid y)\right) \\ &\overset{\text{(b)}}{=} 1 - \sum_{y \in \mathcal{Y}} p(y) \cdot p(\hat{x}_{\mathrm{B}}(y) \mid y), \end{aligned} \tag{8.15}$$

where the equality (a) comes from the fact that the estimate $\hat{x}_{\mathrm{B}}(y)$ provided by the block-MAP decoder is equal to a fixed codeword (more specifically, to the most likely one) with probability 1; and the equality (b) uses that $\sum_{y \in \mathcal{Y}} p(y) = 1$.

Similarly, let $\hat{x}_{\mathrm{B,r}}(y)$ be the estimate provided by the randomized block-MAP decoder, given the channel output y. Then, the following chain of inequalities holds

$$\begin{aligned}
P_{\mathrm{B,r}} &= \sum_{x\in\mathcal{C}} p(x) \sum_{y\in\mathcal{Y}} p(y \mid x)\mathbb{P}(\hat{x}_{\mathrm{B,r}}(y) \neq x) \\
&= \sum_{y\in\mathcal{Y}} p(y) \sum_{x\in\mathcal{C}} p(x \mid y)\mathbb{P}(\hat{x}_{\mathrm{B,r}}(y) \neq x) \\
&\overset{\text{(a)}}{=} \sum_{y\in\mathcal{Y}} p(y) \sum_{x\in\mathcal{C}} p(x \mid y)(1 - p(x \mid y)) \\
&\overset{\text{(b)}}{\leq} \sum_{y\in\mathcal{Y}} p(y) \left(1 - \left(\max_{x\in\mathcal{C}} p(x \mid y)\right)^2\right) \\
&= \sum_{y\in\mathcal{Y}} p(y) \left(1 - p(\hat{x}_{\mathrm{B}}(y) \mid y)^2\right) \\
&\overset{\text{(c)}}{=} 1 - \sum_{y\in\mathcal{Y}} p(y) \cdot p(\hat{x}_{\mathrm{B}}(y) \mid y)^2 \\
&\overset{\text{(d)}}{\leq} 1 - \left(\sum_{y\in\mathcal{Y}} p(y) \cdot p(\hat{x}_{\mathrm{B}}(y) \mid y)\right)^2 \\
&\overset{\text{(e)}}{=} 1 - (1 - P_{\mathrm{B}})^2 \leq 2 \cdot P_{\mathrm{B}}.
\end{aligned} \tag{8.16}$$

To prove the equality (a), we use that the estimate $\hat{x}_{\mathrm{B,r}}(y)$ provided by the randomized block-MAP decoder is equal to x with probability $p(x \mid y)$. To prove inequality (b), we use that, given m real numbers $1 \geq p_1 \geq \cdots \geq p_m \geq 0$ with $\sum_{j=1}^m p_j = 1$, then

$$\begin{aligned}
\sum_{j=1}^m p_j(1 - p_j) &= (1 - p_1) \sum_{j=1}^m p_j \frac{1 - p_j}{1 - p_1} \\
&= (1 - p_1) \left(p_1 + \sum_{j=2}^m p_j \frac{1 - p_j}{1 - p_1}\right) \\
&\leq (1 - p_1) \left(p_1 + \sum_{j=2}^m p_j \frac{1}{1 - p_1}\right) \\
&= (1 - p_1)\,(p_1 + 1) = 1 - p_1^2.
\end{aligned}$$

To prove the equality (c), we use that $\sum_{y\in\mathcal{Y}} p(y) = 1$. Inequality (d) follows from Jensen's inequality and equality (e) uses (8.15).

In order to prove the analogous inequality $P_{\mathrm{b,r}} \leq 2 \cdot P_{\mathrm{b}}$, one possibility is to write expressions similar to (8.15) and (8.16) for the bit error probability of position i under bit-MAP decoding and under randomized bit-MAP decoding, respectively. Otherwise, we can follow the simpler argument used to prove (13.50) in [223]. We reproduce this argument here for the sake of completeness.

Consider first the case of a single bit with posterior probability $\{p_0, p_1\}$. Then, $P_{\mathrm{b}} = \min(p_0, p_1)$. Furthermore, the probability that the randomized bit-MAP de-

coder makes a correct decision is $p_0^2 + p_1^2$, since its output and the ground truth follow the same distribution. Thus,

$$P_{\mathrm{b,r}} = 2p_0p_1 \leq 2\min(p_0, p_1) = 2 \cdot P_{\mathrm{b}}. \tag{8.17}$$

In general, P_{b} and $P_{\mathrm{b,r}}$ are just the averages of many such error probabilities. Therefore, (8.17) holds for the transmission of any number of bits. □

Now, let us state some more definitions and intermediate results that will be useful in the following.

Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a function. The derivative of f in direction $y \in \mathbb{F}_2^n$ is denoted by $\Delta_y f : \mathbb{F}_2^n \to \mathbb{F}_2$ and it is defined as

$$\Delta_y f(x) = f(x+y) + f(x).$$

Similarly, the k-iterated derivative of f in directions $Y = (y_1, \cdots, y_k) \in (\mathbb{F}_2^n)^k$ is denoted by $\Delta_Y f : \mathbb{F}_2^n \to \mathbb{F}_2$ and it is defined as

$$\Delta_Y f(x) = \Delta_{y_1}\Delta_{y_2}\cdots\Delta_{y_k} f(x).$$

A simple manipulation yields

$$\Delta_Y f(x) = \sum_{I \subseteq [k]} f\left(x + \sum_{i \in I} y_i\right),$$

from which it is clear that the order of $y_1, \cdots, y_k$ is irrelevant in the computation of $\Delta_Y f(x)$. Hence, we can think of Y as a multi-set of size k.

Note that, if f is a polynomial of degree v, then its derivatives have degree at most $v-1$. Consequently, its k-iterated derivatives have degree at most $v-k$. Furthermore, as pointed out in Section III of [68], we have that $\Delta_y f(x) = \Delta_y f(x+y)$. Thus, in general, $\Delta_Y f(x)$ is determined by its values on the quotient space $\mathbb{F}_2^n \setminus \langle Y \rangle$, where $\langle Y \rangle$ denotes the space spanned by the vectors in Y.

The following lemma plays a central role in the proof of the upper bound on the weight distribution.

Lemma 8.5 (Lemma 2.1 in [222])**.** *Pick an integer $\ell \geq 1$ and $\epsilon \in (0,1)$. Consider a function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ such that $\mathrm{wt}(f) \leq 2^{-\ell}(1-\epsilon)$. Pick any $\delta > 0$. There exists a universal algorithm $\mathcal{A}$ (which does not depend on f) with the following properties:*

1. *$\mathcal{A}$ has two inputs: $x \in \mathbb{F}_2^n$ and $Y_1, \cdots, Y_t \in (\mathbb{F}_2^n)^\ell$.*
2. *$\mathcal{A}$ has oracle access to the ℓ-iterated derivatives $\Delta_{Y_1} f, \cdots, \Delta_{Y_t} f$.*

Then, for $t \leq C_2(\log_2(1/\delta)\log_2(1/\epsilon) + \log_2(1/\delta)^2)$, where C_2 is a universal constant, there exists a choice of $Y_1, \cdots, Y_t$ such that

$$\mathbb{P}(\mathcal{A}(x; Y_1, \cdots, Y_t, \Delta_{Y_1} f, \cdots, \Delta_{Y_t} f) = f(x)) \geq 1 - \delta, \tag{8.18}$$

where the probability distribution is over $x \in \mathbb{F}_2^n$ chosen uniformly at random.

In words, Lemma 8.5 says that any function of small normalized weight can be approximated arbitrarily well, given a sufficient amount of its derivatives. For a proof of this result, we refer the interested reader to Section II of [222].

At this point, we are ready to prove Lemma 8.4.

Proof of Lemma 8.4. Pick $\delta = 2^{-v-1}$. Apply the universal algorithm $\mathcal{A}$ to all the codewords $f \in \mathrm{RM}(n, v)$. Denote by $\mathcal{H}$ the family of functions obtained by doing so. In other words, $\mathcal{H}$ is the set of outputs of $\mathcal{A}$ when the input is a degree v polynomial in n variables.

By Lemma 8.5, for any $f \in \mathrm{RM}(n, v)$ such that $\mathrm{wt}(f) \leq \alpha$, there exists $h \in \mathcal{H}$ that differs from f in a fraction $< \delta$ of points of $\mathbb{F}_2^n$.

Suppose now that there exists $h \in \mathcal{H}$ obtained by applying the algorithm $\mathcal{A}$ to two distinct codewords $f_1, f_2 \in \mathrm{RM}(n, v)$ such that $\mathrm{wt}(f_1) \leq \alpha$ and $\mathrm{wt}(f_2) \leq \alpha$. Then, h differs from f_1 in a fraction $< \delta$ of points and h differs from f_2 in a fraction $< \delta$ of points. Therefore, f_1 and f_2 can differ in a fraction $< 2\delta = 2^{-v}$ of points. As the minimum distance of the code is 2^{n-v}, we conclude that $f_1 = f_2$. Consequently, we can associate a unique $h \in \mathcal{H}$ with each $f \in \mathrm{RM}(n, v)$ such that $\mathrm{wt}(f) \leq \alpha$. This implies that

$$W_{n,v}(\alpha) \leq |\mathcal{H}|. \tag{8.19}$$

The remainder of the proof consists in upper bounding the cardinality of $\mathcal{H}$.

Recall that the algorithm $\mathcal{A}$ takes as input

1. the t directions $Y_1, \cdots, Y_t \in (\mathbb{F}_2^n)^\ell$ with $t \leq C_2(\log_2(1/\delta)\log_2(1/\epsilon)+\log_2(1/\delta)^2)$;
2. the t ℓ-iterated derivatives of the input.

The number of different possibilities for each Y_i (with $i \in [t]$) is $2^{n\ell}$. Given Y_i, the number of possible functions $\Delta_{Y_i} f$ is upper bounded by the number of polynomials of degree at most $v - \ell$ defined in the space $\mathbb{F}_2^n \setminus \langle Y_i \rangle$. As this space has dimension $n - \ell$, the number of possible functions $\Delta_{Y_i} f$ is $2^{\sum_{j=0}^{v-\ell} \binom{n-\ell}{j}}$.

By putting everything together, we conclude that

$$\begin{aligned}
|\mathcal{H}| &\leq 2^{t(n\ell+\sum_{j=0}^{v-\ell}\binom{n-\ell}{j})} \\
&\overset{(a)}{\leq} 2^{C_2((v+1)\cdot\log_2(1/\epsilon)+(v+1)^2)(n\ell+\sum_{j=0}^{v-\ell}\binom{n-\ell}{j})} \\
&\overset{(b)}{\leq} 2^{C_2((v+1)\cdot\log_2(1/\epsilon)+(v+1)^2\cdot\log_2(1/\epsilon))(n\ell+\sum_{j=0}^{v-\ell}\binom{n-\ell}{j})} \\
&\leq (1/\epsilon)^{C_2(v+1)(v+2)(n\ell+\sum_{j=0}^{v-\ell}\binom{n-\ell}{j})} \\
&\leq (1/\epsilon)^{C_2(v+2)^2(n\ell+\sum_{j=0}^{v-\ell}\binom{n-\ell}{j})},
\end{aligned}$$

where the inequality (a) combines the upper bound on t with the choice $\delta = 2^{-v-1}$; and the inequality (b) uses that $\log_2(1/\epsilon) \geq 1$ for $\epsilon \in (0, 1/2]$. □

As previously pointed out, Lemma 8.4 is a refinement of Theorem 3.1 of [222]. More specifically, the upper bound (8.19) comes from the proof of Theorem 3.1 of [222], and our refinement consists in an improved upper bound on $|\mathcal{H}|$. Note that this improvement is necessary to obtain the desired result on the error probability of

the randomized block-MAP decoder, as the upper bound on the weight distribution presented in Theorem 3.1 of [222] is not tight enough for this purpose.

Let us proceed with the proof of Lemma 8.3.

Proof of Lemma 8.3. Let $\{\mathrm{RM}(n, v_n)\}$ be a sequence of Reed-Muller codes with block lengths $N_n = 2^n \to \infty$ and rates $R_n \to R$, for $R \in (0, 1)$. Since the claim to be proved is stronger when β is smaller, we can assume without loss of generality that $\beta \in (0, 1/2)$.

Suppose now that, for n large enough,

$$v_n > \frac{n}{2}\left(1 + \frac{\beta}{2}\right). \tag{8.20}$$

Then, by applying Lemma 4.7.2 of [217], we obtain that

$$R_n \geq 1 - \frac{2^{nh_2(\frac{n-v_n-1}{n})}}{2^n},$$

as $n - v_n - 1 \leq n/2$. This means that, for any $\beta \in (0, 1/2)$, if v_n satisfies (8.20), then the rate R_n tends to 1.

Similarly, it is easy to see that if

$$v_n < \frac{n}{2}\left(1 - \frac{\beta}{2}\right),$$

then the rate R_n tends to 0. Since R_n converges to $R \in (0, 1)$, we have that, for n large enough,

$$v_n \in \left(\frac{n}{2}\left(1 - \frac{\beta}{2}\right), \frac{n}{2}\left(1 + \frac{\beta}{2}\right)\right). \tag{8.21}$$

Let x denote a codeword and y the channel output. Then, the posterior probability $p(x \mid y)$ can be written as

$$p(x \mid y) = \frac{p(y \mid x)p(x)}{\sum_{\tilde{x}} p(y \mid \tilde{x})p(\tilde{x})} = \frac{p(y \mid x)}{\sum_{\tilde{x}} p(y \mid \tilde{x})}, \tag{8.22}$$

where the last equality comes from the fact that the codeword is chosen uniformly from the codebook. From (8.22) we deduce that, by adding codewords, the posterior probability $p(x \mid y)$ decreases. Then, the probability that the randomized block-MAP decoder outputs a specific codeword x increases if we remove all codewords except x and the codeword that was actually transmitted. By using (8.12), we can upper bound such a probability by 2 times the block error probability of the non-randomized block-MAP decoder. Eventually, by applying Lemma 4.67 of [44], this last probability is upper bounded by $\frac{1}{2}z^w$, where w is the Hamming weight of x.

The argument above proves that the probability that the randomized block-MAP decoder outputs a codeword of weight $w \in [2^n]$ is upper bounded by z^w. Hence, by applying the union bound, the probability that the randomized block-MAP decoder outputs a codeword of normalized weight at most $2^{-n\beta}$ is upper bounded by

$$\sum_{w=1}^{\lceil 2^{n(1-\beta)} \rceil} z^w c_w,$$

where c_w denotes the number of codewords of weight w. As the minimum distance of the code $\mathrm{RM}(n, v_n)$ is 2^{n-v_n}, we deduce that $c_w = 0$ for $w \in \{1, \cdots, 2^{n-v_n} - 1\}$. For $w \in \{2^{n-v_n}, \cdots, \lceil 2^{n(1-\beta)} \rceil\}$, we have that

$$\begin{aligned}
\log_2(c_w) &\overset{(a)}{\leq} \log_2\left(W_{n,v_n}(w2^{-n})\right) \\
&\overset{(b)}{\leq} \log_2\left(W_{n,v_n}(2^{\lceil \log_2(w) \rceil - n})\right) \\
&\overset{(c)}{\leq} C_2(v_n+2)^2\left(n(n-\lceil \log_2(w) \rceil - 1) + \sum_{i=0}^{v_n - n + \lceil \log_2(w) \rceil + 1} \binom{\lceil \log_2(w) \rceil + 1}{i}\right) \\
&\overset{(d)}{\leq} C_2\, n^2\left(n^2 + 2^{(\lceil \log_2(w) \rceil + 1)\cdot h_2\left(\frac{v_n - n + \lceil \log_2(w) \rceil + 1}{\lceil \log_2(w) \rceil + 1}\right)}\right) \\
&\overset{(e)}{\leq} C_2\, n^2\left(n^2 + 2^{(\log_2(w)+2)\cdot h_2\left(\frac{n\beta/4 - n/2 + \log_2(w) + 2}{\log_2(w)}\right)}\right),
\end{aligned} \tag{8.23}$$

where the inequality (a) comes from the definition (8.13) of cumulative weight distribution; the inequality (b) comes from the fact that $W_{n,v_n}(\alpha)$ is increasing in α; the inequality (c) comes from the application of Lemma 8.4 with $\ell = n - \lceil \log_2(w) \rceil - 1$ and $\epsilon = 1/2$; the inequality (d) comes from the application of Lemma 4.7.2 of [217] (or, equivalently, of formula (1.59) of [44]); and the inequality (e) comes from the fact that $h_2(x)$ is increasing for $x \in [0, 1/2]$ and v_n is upper bounded by (8.21). Note that we fulfill the hypotheses of Lemma 8.4 since $w \geq 2^{n-v_n}$ implies that $\ell \leq v_n - 1$. In addition, we can apply Lemma 4.7.2 of [217] since (8.21) and $w \leq \lceil 2^{n(1-\beta)} \rceil$ imply that $v_n - n + \lceil \log_2(w) \rceil + 1 \leq (\lceil \log_2(w) \rceil + 1)/2$ for n large enough.

Thus, the logarithm of the desired probability is upper bounded as follows:

$$\begin{aligned}
\log_2\left(\sum_{w=1}^{\lceil 2^{n(1-\beta)} \rceil} z^w c_w\right) &\overset{(a)}{=} \log_2\left(\sum_{w=2^{n-v_n}}^{\lceil 2^{n(1-\beta)} \rceil} z^w c_w\right) \\
&\overset{(b)}{\leq} n + \max_{w \in \mathbb{N} \cap [2^{n-v_n}, \lceil 2^{n(1-\beta)} \rceil]} \log_2(z^w c_w) \leq n + \max_{w \in [2^{n-v_n}, \lceil 2^{n(1-\beta)} \rceil]} \log_2(z^w c_w) \\
&\overset{(c)}{\leq} n + C_2\, n^4 + \max_{\log_2(w) \in [n - v_n, n(1-\beta)+1]} \Bigg(- \log_2(1/z) \cdot 2^{\log_2(w)} \\
&\qquad + C_2\, n^2 2^{(\log_2(w)+2)\cdot h_2\left(\frac{n\beta/4 - n/2 + \log_2(w) + 2}{\log_2(w)}\right)} \Bigg) \\
&\overset{(d)}{\leq} n + C_2\, n^4 + \max_{x \in [1/2 - \beta/4, (1-\beta) + 1/n]} \Bigg(- \log_2(1/z) \cdot 2^{nx} \\
&\qquad + C_2\, n^2 2^{n(x+2/n)\cdot h_2\left(\frac{\beta/4 - 1/2 + x + 2/n}{x}\right)} \Bigg) \\
&\overset{(e)}{\leq} n + C_2\, n^4 + \max_{x \in [1/2 - \beta/4, 1 - 7\beta/8]} \left(- \log_2(1/z) \cdot 2^{nx} + 4C_2\, n^2 2^{nx \cdot h_2\left(\frac{\beta/3 - 1/2 + x}{x}\right)} \right),
\end{aligned} \tag{8.24}$$

where the equality (a) uses that $c_w = 0$ for $w \in \{1, \cdots, 2^{n-v_n} - 1\}$; the inequality (b) uses that the number of terms in the sum is upper bounded by 2^n; the inequality (c) uses (8.23); to obtain the inequality (d), we set $x = \log_2(w)/n$ and we use the upper bound (8.21) on v_n; and the inequality (e) uses that

$$\begin{aligned} h_2(t) &\leq 1, \\ 1 - \frac{7\beta}{8} &\geq 1 - \beta + \frac{1}{n}, \\ h_2\left(\frac{\beta/4 - 1/2 + x + 2/n}{x}\right) &\leq h_2\left(\frac{\beta/3 - 1/2 + x}{x}\right), \end{aligned}$$

where the last two inequalities hold for n large enough.

In order to conclude, it suffices to observe that, for any $\beta \in (0, 1/2)$ and any $x \in [1/2 - \beta/4, 1 - 7\beta/8]$, we have

$$h_2\left(\frac{\beta/3 - 1/2 + x}{x}\right) < 1,$$

which implies that the upper bound in (8.24) tends to $-\infty$, hence the desired probability goes to 0. □

The following two remarks discuss how to tighten the main result by making the hypothesis on the decay rate of P_{b} less restrictive and by evaluating the decay rate of P_{B}.

Remark 8.3 (Looser Condition on P_{b})**.** *Consider a sequence of Reed-Muller codes $\{RM(n, v_n)\}$ with block lengths $N_n = 2^n \to \infty$ and rates $R_n \to R$, for $R \in (0, 1)$. In order to have that $P_{\mathrm{B}} \to 0$, Theorem 8.5 requires that P_{b} is $O(N_n^{-\gamma}) = O(2^{-n\gamma})$ for some $\gamma > 0$. With some more work, we can conclude that $P_{\mathrm{B}} \to 0$ even under the less restrictive hypothesis that P_{b} is $O(2^{-n^{1/2+\gamma'}})$ for some $\gamma' > 0$. The proof of this tighter result is based on a stronger version of Lemma 8.3 that is outlined in the next paragraph.*

Consider the same transmission scenario of Lemma 8.3 and fix any $\beta' > 0$. Then, the probability that the randomized block-MAP decoder outputs an incorrect codeword at Hamming distance at most $2^{n-n^{1/2+\beta'}}$ tends to 0 as n tends to infinity. In other words, codewords with distances up to $2^{n-n^{1/2+\beta'}}$, for any $\beta' > 0$, do not count, as opposed to distances up to $2^{n(1-\beta)}$, for any $\beta > 0$, in the original statement. In order to prove this stronger claim, first we need the following tighter bound for the range of v_n (compare to (8.21)):

$$v_n \in \left(\frac{n}{2} - \frac{n^{1/2+\beta'}}{4}, \frac{n}{2} + \frac{n^{1/2+\beta'}}{4}\right).$$

Indeed, from simple manipulations we have that $v_n > n/2 + n^{1/2+\beta'}/4$ yields rates $R_n \to 1$ and $v_n < n/2 - n^{1/2+\beta'}/4$ yields rates $R_n \to 0$. Then, we obtain this bound on $\log_2(c_w)$ (compare to the last inequality in (8.23)):

$$\log_2(c_w) \leq C_2\, n^2 \left(n^2 + 2^{(\log_2(w)+2) h_2\left(\frac{\frac{n^{1/2+\beta'}}{4} - \frac{n}{2} + \log_2(w) + 2}{\log_2(w)}\right)} \right),$$

which yields the following upper bound on the logarithm of the desired probability (compare to the last inequality in (8.24)*),*

$$n + C_2\, n^4 + \max_{x\in[1/2-n^{\beta'-1/2}/4,\,1-7n^{\beta'-1/2}/8]} \left(-\log_2(1/z)\cdot 2^{nx} + 4\, C_2\, n^2 2^{nx\cdot h_2\left(\frac{n^{\beta'-1/2}/3-1/2+x}{x}\right)} \right). \tag{8.25}$$

Eventually, when $n \to \infty$, we can show that the above quantity tends to $-\infty$, which suffices to prove the claim. Note that the result of Theorem 8.5 cannot be further improved by using a better upper bound on the weight distribution. Indeed, a simple counting argument gives that $W_{n,v}(2^{-\ell}) \geq 2^{n\ell+\binom{n-\ell}{v-\ell}}$ (see also Section III of [68]).

Remark 8.4 (Decay Rate of P_{B})**.** *The decay rate of P_{B} is given by the slowest between the decay rates of $P_{\mathrm{B,r}}^{\mathrm{l}}$ and $P_{\mathrm{B,r}}^{\mathrm{h}}$, defined in the proof of Theorem 8.5.*

First, assume that P_{b} is $O(2^{-n\gamma})$, for some $\gamma > 0$, as in the hypothesis of the theorem. Note that (8.24) *is minimized when $x = 1/2 - \beta/4$ and we can pick any $\beta \leq \gamma/2$, since β is set to $\gamma/2$ in the proof of Theorem 8.5 and the claim of Lemma 8.3 is stronger when β is smaller. Therefore, we obtain that $P_{\mathrm{B,r}}^{\mathrm{l}}$ is $O(2^{-2^{n\tau}})$, for any $\tau \in (0, 1/2)$. This bound essentially comes from the fact that the minimum distance of Reed-Muller codes scales as $\sqrt{N}$, with $N = 2^n$. From the argument in the last paragraph of the proof of Theorem 8.5, we have that $P_{\mathrm{B,r}}^{\mathrm{h}}$ is $O(2^{-n\rho})$, for any $\rho \in (0, \gamma)$. Thus, when P_{b} is $O(2^{-n\gamma})$, we conclude that $P_{\mathrm{B,r}}^{\mathrm{h}}$ is $O(2^{-n\rho})$ for any fixed $\rho < \gamma$.*

Now, assume that P_{b} is $O(2^{-n^{1/2+\gamma'}})$, for some $\gamma' > 0$, as in Remark 8.3. From (8.25)*, we obtain again that $P_{\mathrm{B,r}}^{\mathrm{l}}$ is $O(2^{-2^{n\tau}})$, for any $\tau \in (0, 1/2)$. From the argument in the proof of Theorem 8.5, we have that $P_{\mathrm{B,r}}^{\mathrm{h}}$ is $O(2^{-a\cdot n^{1/2+\gamma'}})$, for any $a \in (0, 1)$, which also gives the overall decay rate of P_{B}. In conclusion, these arguments show that the decay rates of P_{b} and P_{B} are essentially the same.*

8.6 Appendix

8.6.1 Proof of Theorem 8.4

Proof. By applying Theorem 1 and Corollary 4.1 of [218], we have that there exists a universal constant C_1' such that

$$\frac{d\mu_\varepsilon(\Omega)}{d\varepsilon} \geq C_1' \ln(\ln M) \ln(M) \mu_\varepsilon(\Omega)(1 - \mu_\varepsilon(\Omega)), \tag{8.26}$$

provided that $\varepsilon(1-\varepsilon)$ is bounded away from 0.

Define

$$g(\varepsilon) = \ln \frac{\mu_\varepsilon(\Omega)}{1 - \mu_\varepsilon(\Omega)}.$$

Then,

$$\frac{dg(\varepsilon)}{d\varepsilon} = \frac{1}{\mu_\varepsilon(\Omega)(1-\mu_\varepsilon(\Omega))} \frac{d\mu_\varepsilon(\Omega)}{d\varepsilon} \geq C_1' \ln\ln M \ln(M), \tag{8.27}$$

where the last inequality comes from (8.26) and we assume that $\varepsilon(1-\varepsilon)$ is bounded away from 0.

By integrating $dg(\varepsilon)/d\varepsilon$ from $\varepsilon^*(\Omega,\delta)$ to $\varepsilon^*(\Omega,1-\delta)$, we obtain that

$$\begin{aligned}\int_{\varepsilon^*(\Omega,\delta)}^{\varepsilon^*(\Omega,1-\delta)} \frac{dg(\varepsilon)}{d\varepsilon}\, d\varepsilon &= g(\varepsilon^*(\Omega,1-\delta)) - g(\varepsilon^*(\Omega,\delta)) \\ &= \ln\frac{1-\delta}{\delta} - \ln\frac{\delta}{1-\delta} \\ &= 2\ln\frac{1-\delta}{\delta} \leq 2\ln\frac{1}{\delta}.\end{aligned}$$

Furthermore, the application of (8.27) yields that

$$\int_{\varepsilon^*(\Omega,\delta)}^{\varepsilon^*(\Omega,1-\delta)} \frac{dg(\varepsilon)}{d\varepsilon}\, d\varepsilon \geq C_1' \ln\ln M \ln(M)\cdot(\varepsilon^*(\Omega,1-\delta) - \varepsilon^*(\Omega,\delta)).$$

Hence, the desired result follows by setting $C_1 = 1/(2C_1')$. □

8.6.2 Proof of Lemma 8.1

Proof. As in the proof of Lemma 7.1 in Appendix 7.6.1, the elements of the vector space $\{0,1\}^n$ are enumerated by $e^{(1)},\cdots,e^{(N)}$, where $N=2^n$ and $e^{(N)}$ is associated with the sequence of n 0s, namely, $e^{(N)} = 0_{1:n}$. Recall also that $P(n,v)$ denotes the set of multivariate polynomial with n binary variables of degree at most v and that the codewords of the code $\mathrm{RM}(n,v)$ are of the form $(f(e^{(1)}),\cdots,f(e^{(N)}))$, with $f\in P(n,v)$.

Let us associate with a given $T\in \mathrm{GL}(n,\mathbb{F}_2)$ the permutation $\pi_T\in S_{N-1}$ such that

$$\pi_T(\ell) = \ell', \quad \text{where } e^{(\ell')} = Te^{(\ell)}.$$

Note that π_T is well-defined since T is invertible. Moreover, it is easy to check that $\pi_{T_1}\circ\pi_{T_2} = \pi_{T_1T_2}$ for $T_1,T_2\in\mathrm{GL}(n,\mathbb{F}_2)$. As such, the collection of permutations

$$\mathcal{H} = \{\pi_T\in S_{N-1} \mid T\in \mathrm{GL}(n,\mathbb{F}_2)\}$$

is a subgroup of S_{N-1} isomorphic to $\mathrm{GL}(n,\mathbb{F}_2)$.

For $i,j\in[N-1]$, there exists $T\in\mathrm{GL}(n,\mathbb{F}_2)$ such that $e^{(j)} = Te^{(i)}$ and, thus, $\pi_T(i) = j$. Consequently, $\mathcal{H}$ is transitive.

In order to finish the proof, it remains to show that $\mathcal{H}\subseteq\mathcal{G}_N$. To do so, associate $\pi_T\in\mathcal{H}$ with $\pi_T'\in S_N$ where

$$\pi_T'(\ell) = \pi_T(\ell) \quad \text{for } \ell\in[N-1], \qquad\qquad \pi_T'(N) = N.$$

Let us prove that π_T' belongs to the permutation group $\mathcal{G}$ of the code $\mathrm{RM}(n,v)$. Consider a codeword given by $f\in P(n,v)$ and let g be defined as

$$g(x_1,\cdots,x_n) = f(T^{-1}[x_1,\cdots,x_n]^{\mathrm{T}}).$$

Then, $\mathrm{degree}(g) = \mathrm{degree}(f)$ and $g(e^{(\pi_T'(\ell))}) = f(e^{(\ell)})$ for $\ell\in[N-1]$, as $e^{(\ell')} = Te^{(\ell)}$. Furthermore, $g(e^{(N)}) = f(T^{-1}0_{1:n}) = f(e^{(N)})$, which implies that $\pi_T'\in\mathcal{G}$.

Since $\mathcal{G}_N$ is the permutation group of the set Ω_N of Definition 7.1, it is clear that if $\pi_T'\in\mathcal{G}$, then $\pi_T\in\mathcal{G}_N$, which concludes the proof. □

9 Conclusions and Perspectives

Una frase compiuta deve avere.

A complete sentence must comprise.

In this concluding chapter, we summarize our most important findings, we highlight that the methods and proof techniques developed so far can also be helpful for other problems, and we describe some future research directions. In particular, each of the first three sections covers one of the three main themes of this work, i.e., unified scaling, non-standard channels, and capacity via symmetry.

At the beginning of the thesis, after a brief historical introduction, we addressed directly the reader and asked "What Now?". As we talked extensively about symmetry, it seems appropriate to conclude by asking another question that is, "What's Next?".

9.1 Unified Scaling

In **Chapter 2**, we present a *unified view* on the *scaling of polar codes* by studying the relationship of the fundamental parameters at play, i.e., the block length N, the rate R, the block error probability under successive cancellation decoding P_{B}, the capacity of the transmission channel $C(W)$ and its Bhattacharyya parameter $Z(W)$.

First of all, we prove a new *upper bound* on the *scaling exponent* valid for any BMS channel W. The setting is the following: we fix the error probability P_{B} and we study how the gap to capacity $C(W) - R$ scales with the block length N. In particular, N is $O\left(1/(C(W) - R)^{\mu}\right)$, where μ is the so-called scaling exponent whose value depends on W, and we show a better upper bound on μ valid for any BMS channel W. The proof technique consists in relating the value of μ to the supremum of a function that fulfills certain constraints. Then, we upper bound the supremum by constructing and analyzing a suitable candidate function. We underline that the proposed bound is *provable* and that the analysis of the algorithm is not affected by numerical errors, as all the computations can be reduced to computations over integers, thus they can be performed exactly. The proposed proof technique yields

$\mu \leq 4.714$ for any BMS channel, which improves by 1 the existing upper bound. If W is a BEC, then we obtain $\mu \leq 3.639$, which closely approaches the value previously computed with heuristic methods. These bounds can be slightly tightened simply by increasing the number of samples used by the algorithm.

Second, we consider a *moderate deviations* regime and we prove a trade-off between the speed of decay of the error probability and that of the gap to capacity. The setting is the following: we do not fix either the error probability P_{B} or the gap to capacity $C(W) - R$, but we study how fast both P_{B} and $C(W) - R$, as functions of the block length N, go to 0 at the same time. In particular, we show that, if the gap to capacity is such that

$$N = O\left(\frac{1}{(C(W) - R)^{\mu/(1-\gamma)}}\right), \qquad \text{for } \gamma \in \left(\frac{1}{(1+\mu)}, 1\right),$$

then the error probability is given by

$$P_{\mathrm{B}} = O\left(N \cdot 2^{-N^{\gamma \cdot h_2^{(-1)}\left(\frac{\gamma(\mu+1)-1}{\gamma\mu}\right)}}\right).$$

Note that, as the exponents $\mu/(1-\gamma)$ and $\gamma \cdot h_2^{(-1)}\left((\gamma(\mu+1)-1)/(\gamma\mu)\right)$ are both increasing in γ, if the error probability decays faster, then the gap to capacity decays slower. This trade-off recovers the existing result for the error exponent regime, but it does not match the new bound on the scaling exponent. An interesting open question consists in finding the optimal trade-off that would provide the fastest possible decay of the error probability, given a certain speed of decay of the gap to capacity. Note that this optimal trade-off would match the existing results for both the error exponent and the scaling exponent regimes.

Third, we prove that polar codes are *not affected by error floors*. The setting is the following: we fix a polar code of block length N and rate R designed for a channel W', we let the transmission channel W vary, and we study how the error probability $P_{\mathrm{B}}(W)$ scales with the Bhattacharyya parameter $Z(W)$ of the channel W. In particular, we show that

$$P_{\mathrm{B}}(W) \leq Z(W)^{\frac{\log_2 \tilde{P}_{\mathrm{B}}(W')}{\log_2 Z(W')}},$$

where $\tilde{P}_{\mathrm{B}}(W')$ denotes the sum of the Bhattacharyya parameters at the information positions obtained by polarizing W'. In addition, $\log_2 \tilde{P}_{\mathrm{B}}(W')/\log_2 Z(W')$ scales roughly as $\sqrt{N}$ and this is the best possible scaling according to the error exponent regime. Hence, the scaling between P_{B} and $Z(W)$ would have been the same, even if we "matched" the code to the channel. However, when W and W' can be any BMS channel, the result holds only if $Z(W) \leq Z(W')^2$. An interesting open question is to explore further the case $Z(W) \in (Z(W')^2, Z(W')]$, in order to see whether a similar but perhaps less tight bound still holds.

Let us highlight that the *technical tools* developed in this chapter have proved *useful* also in different scenarios. Indeed, the analysis of Section 2.3 is the starting point for the characterization of the scaling exponent of binary-input energy-harvesting channels [224] and of q-ary polar codes based on $q \times q$ Reed-Solomon polarization kernels [225].

Why are we interested in $q \times q$ kernels? Such kernels have the potential to improve the scaling behavior of polar codes. For the error exponent, in [226] it is proved that, as q goes large, the error probability scales roughly as 2^{-N}. For the scaling exponent, in [227] it is observed that μ can be reduced when $q \geq 8$. In the recent paper [225], it is shown that, for the transmission over the erasure channel, the optimal scaling exponent $\mu = 2$ is approached by using a large kernel and a large alphabet. Furthermore, in [43], the author gives evidence supporting the conjecture that, in order to obtain $\mu = 2$, it suffices to consider a large random kernel over a binary alphabet. Hence, providing a rigorous proof of such a conjecture is a very interesting open problem.

Another approach to improving the scaling exponent consists in acting on the decoding algorithm. In particular, the successive cancellation list decoder proposed in [55] provides a significant performance improvement. However, in **Chapter 3**, we present a negative result: the introduction of any *finite list cannot improve* the *scaling exponent* under MAP decoding. The proof technique is based on a Divide and Intersect (DI) procedure that lower bounds the error probability under MAP decoding with list size L for any BMS channel. The result that we obtain is very general, as it applies not only to polar codes but to any family of linear codes with an unbounded minimum distance.

A similar DI bound is proved for the genie-aided successive cancellation decoder, when the transmission takes place over the BEC. Consequently, the *scaling exponent* under genie-aided decoding *does not change* for any *fixed number of helps from the genie*. Note that, as genie-aided SC decoding might be strictly worse than successive cancellation list decoding, the problem of establishing the scaling exponent of the latter remains open.

These results suggest that an improvement only in the decoding algorithm might not be enough to change the scaling exponent. Hence, in **Chapter 6**, we address the issue of *boosting the finite-length performance* of polar codes by *modifying jointly the code and the decoding* algorithm. In particular, we construct the family of codes $\{\mathcal{C}_\alpha\}$, for $\alpha \in [0, 1]$, of fixed block length and rate; this family interpolates from the original polar code $\mathcal{C}_\alpha\big|_{\alpha=1}$ to the Reed-Muller code $\mathcal{C}_\alpha\big|_{\alpha=0}$. Numerically, the error probability under MAP decoding decreases as α goes from 1 to 0. As MAP decoding is not practical for the transmission over general channels, we develop a trade-off between complexity and performance by considering low-complexity decoders (e.g., belief propagation, list decoding). As a result, we show the significant benefit coming from the adoption of codes from the family $\{\mathcal{C}_\alpha\}$ via numerical simulations for the BEC and the binary Gaussian channel. Note that this performance improvement comes at no additional cost: $\mathcal{C}_\alpha$ is simply a polar code for a mismatched channel, hence the encoding and decoding algorithms are the same as for the original polar code $\mathcal{C}_\alpha\big|_{\alpha=1}$. In addition, we provide experimental evidence of the fact that the error probability under MAP decoding for the transmission over the BEC of $\mathcal{C}_\alpha$ for α sufficiently small is very close to the error probability of random codes. As random codes achieve the optimal scaling exponent $\mu = 2$, the family $\{\mathcal{C}_\alpha\}$ has the potential to improve the scaling behavior of polar codes.

9.2 Non-standard Channels

In **Chapter 4**, we consider the two-user discrete memoryless broadcast channel and we show how to construct polar codes that achieve the superposition and binning regions. By combining these two strategies, we *achieve* any rate pair inside *Marton's region* with both common and private messages. This rate region is tight for all classes of broadcast channels whose capacities are known and, in general, it constitutes the *best existing inner bound*. The described coding techniques possess the usual advantages of polar codes, i.e., encoding and decoding complexity of $\Theta(N \log_2 N)$ and block error probability decaying like $O(2^{-N^\beta})$ for any $\beta \in (0, 1/2)$.

The current exposition is limited to the case of binary auxiliary random variables and, for Bergmans' superposition coding scheme, also to binary inputs. However, there is no fundamental difficulty in extending our schemes to the q-ary case, building on the existing polar constructions for channels with arbitrary input alphabets [79–84]. It is also easy to extend the proposed polar coding techniques to obtain inner bounds for the K-user broadcast channel in a low-complexity fashion.

It is worth pointing out that the chaining construction used to align the polarized indices does not rely on the specific structure of the broadcast channel. Indeed, this method was later used to design polar coding schemes for many other communication settings, e.g., noisy write-once memories [116], general wiretap channels [117], broadcast and wiretap channels with confidential messages [118, 119]. Actually, the *chaining construction is a general coding primitive* whose applicability is not restricted to polar codes, as we demonstrate in the following chapter.

In particular, in **Chapter 5**, we consider another non-standard setting and survey *three paradigms for achieving the capacity of asymmetric channels*.

The first approach is based on *Gallager's mapping*. The idea was first described in [120], and it consists of employing a non-linear function in order to make the input distribution match the capacity-achieving one. In this way, we can achieve the capacity of asymmetric channels by using either q-ary or binary codes that are capacity-achieving for suitably defined symmetric channels.

The second approach consists in an *integrated scheme* that simultaneously performs the tasks of source coding and of channel coding. The idea was first presented for polar codes in [121], and here we extend it to sparse graph codes. Indeed, sparse graph codes can be effectively used to create biased codewords from uniform bits (source coding part) and to provide error correction (channel coding part). Given the vector of syndromes, we generate the codeword by running a belief-propagation algorithm with decimation steps. This technique works well in practice, but the proof that the scheme is capacity-achieving remains an open problem.

The third approach consists in a *chaining construction*, where we consider the transmission of k blocks and use a part of the current block to store the syndromes coming from the previous block. The idea was first proposed in [122], and here we show how to use it to provably achieve the capacity of asymmetric channels. By decoupling completely the source coding from the channel coding task, we can employ an optimal scheme to reach each of these two objectives separately. Thus, many combinations are possible: for example, we can use polar codes or arithmetic codes for the source coding part, and polar codes or spatially coupled codes for the channel coding part.

As for the integrated scheme and the chaining construction, we restrict our

discussion to the case of binary-input channels. In order to extend our results to channels with an arbitrary finite-input alphabet, we require schemes that solve the source coding and the channel coding tasks in the non-binary case. For the source coding part, several papers have focused on the construction of polar codes for arbitrary input alphabets [79–84]. Furthermore, we can also easily generalize the solution based on arithmetic coding to non-binary alphabets. For the channel coding part, recall that in Section 5.3 we have converted a non-binary channel into several binary channels by using the chain rule of mutual information (see formula (5.13)). Here, the same idea can be applied as well. Alternatively, we can use directly non-binary spatially coupled codes [87–90] or non-binary polar codes.

9.3 Capacity via Symmetry

In **Chapter 7**, we prove that binary linear *codes with doubly transitive permutation groups achieve capacity* over the BEC for any rate $R \in (0, 1)$ under bit-MAP decoding. Consequently, we are able to show that *Reed-Muller codes are capacity-achieving*, thus settling a long-standing conjecture.

By taking advantage of the symmetry of the code, we obtain that the extrinsic information transfer (EXIT) functions associated with the various positions are all equal to the average EXIT function, and they can be written as the measure of monotone and symmetric sets. According to an important result in theoretical computer science, the functions of this type experience a sharp threshold. Ultimately, this approach is successful because the transition point of the average EXIT function, closely related to the error probability under bit-MAP decoding, is known a priori from the area theorem. One remarkable aspect of our method consists in its simplicity. In particular, we do not use in any way the precise structure of the codes. This means that *symmetry alone* suffices to give *optimal performance*.

One natural question is to what extent the hypothesis of double transitivity can be relaxed, while keeping the capacity-achieving property. Some progress on this point is provided by the recent work [228], where it is proved that a large family of cyclic codes, whose permutation groups satisfy a condition weaker than double transitivity, achieves capacity on erasure channels. The extension of these ideas to the quantum erasure channel is provided in [229], where it is proved that Reed-Muller codes are capacity-achieving also in this scenario.

In **Chapter 8**, we discuss several other *generalizations*. In particular, we consider the case of rates $R \to 0$ and $R \to 1$, and we present two methods for extending results on the bit-MAP error probability to the block-MAP error probability. For the first topic, we prove that Reed-Muller codes are capacity-achieving in a new regime that does not overlap with the one previously considered in [68, 69]. For the second topic, we start by presenting the proof in [192] that Reed-Muller codes achieve capacity under block-MAP decoding. This approach relies on the sharp threshold framework developed in [218] and, as such, it applies only to the case of the transmission over an erasure channel. The main contribution of the chapter consists in the presentation of a *general method for strengthening results on the bit-MAP threshold to the block-MAP threshold* via the careful analysis of the *weight distribution* of Reed-Muller codes. In particular, we show that, if the error probability under bit-MAP decoding tends to 0 with sufficient speed, then the error probability

under block-MAP decoding also tends to 0. This result applies to the transmission over any BMS channel. Therefore, it can be considered as a first step towards the generalization of the ideas in [192, 193] beyond the erasure channel.

Indeed, one of the main open questions consists in showing that codes with sufficient symmetry and, more specifically, Reed-Muller codes are capacity-achieving for the transmission over any memoryless symmetric channel. In such a general setting, the EXIT function can be replaced by the generalized EXIT (GEXIT) function [214]. The area theorem still holds but, in order to apply the sharp transition framework, new ideas will certainly be required, because the straightforward approach leads to the analysis of functions that cannot be written as measures of monotone sets.

9.4 What's Next?

We conclude our journey from polar to Reed-Muller codes by suggesting three research directions, one for each of the three main themes that are the basis of this work.

In regard to *scaling*, the behavior of polar codes under successive cancellation decoding is well understood. In order to boost their performance, some improvements to the decoding algorithm and to the code have also been proposed. However, none of these improvements has led to a coding scheme with a provably better scaling exponent. Hence, the challenge is to **design a polar-like code** that shows a **significant performance gain** at the block lengths typically considered in applications and that achieves the **optimal scaling exponent** $\mu = 2$.

In regard to *non-standard channels*, many of the rate regions previously obtained with random coding arguments have been recently achieved in a low-complexity fashion. To do so, new coding primitives have been devised and these techniques are quite different from the classic schemes of network information theory. For example, in this thesis we have exploited the so-called chaining construction. However, to the best of the author's knowledge, all these low-complexity schemes can only transmit at rates that were already known to be information-theoretically achievable. Hence, the challenge is to exploit the novelty of these building blocks designed specifically for a **practical coding system**, in order to **achieve a new and tighter inner bound** on the capacity region of a multi-user scenario.

In regard to *capacity via symmetry*, we have discovered a new paradigm to achieve capacity and we have already mentioned that an interesting open direction is to generalize our approach to the transmission over any BMS channel. In addition, let us point that our results hold under optimal MAP decoding. For the case of the erasure channel, MAP decoding is equivalent to the inversion of a linear system and, as such, can be performed in polynomial time. However, for general channels, this task has exponential complexity and, for this reason, it is impractical. Hence, the challenge is to **find a low-complexity decoding algorithm for Reed-Muller codes** with near-optimal performance.

As a final note, recall that, in Chapter 6, we have pointed out how the performance of Reed-Muller codes over the BEC under MAP decoding is close to that of random codes. As random codes achieve the optimal scaling exponent, the optimist would say that the same argument can solve both the first and the third problem stated above.

Bibliography

[1] S. B. Wicker and V. K. Bhargava, *Reed-Solomon Codes and their Applications*. IEEE Press, 1994.

[2] P. Elias, "Error-free coding," *IEEE Trans. Inform. Theory*, vol. 4, pp. 29–37, Sep. 1954.

[3] C. E. Shannon, "A mathematical theory of communication," *Bell System Tech. J.*, vol. 27, pp. 379–423, 623–656, July and Oct. 1948.

[4] A. Feinstein, "A new basic theorem of information theory," *IRE Trans. Inform. Theory*, vol. 4, no. 4, pp. 2–22, Sept. 1954.

[5] P. Elias, "Coding for noisy channels," in *IRE International Convention Record*, Mar. 1955, pp. 37–46.

[6] J. Wolfowitz, "The coding of message subject to chance errors," *Illinois J. Math.*, vol. 1, no. 4, pp. 591–606, 1957.

[7] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inform. Theory*, vol. 11, no. 1, pp. 3–18, Jan. 1965.

[8] G. Poltyrev, "Bounds on the decoding error probability of binary linear codes via their spectra," *IEEE Trans. Inform. Theory*, vol. 40, no. 4, pp. 1284–1292, July 1994.

[9] N. Shulman and M. Feder, "Random coding techniques for nonrandom codes," *IEEE Trans. Inform. Theory*, vol. 45, no. 6, pp. 2101–2104, Sept. 1999.

[10] D. J. Costello Jr. and G. D. Forney Jr., "Channel coding: The road to channel capacity," *Proc. of the IEEE*, vol. 95, no. 6, June 2007.

[11] E. R. Berlekamp, *Algebraic Coding Theory*. New York: McGraw-Hill, 1968.

[12] F. J. MacWilliams and N. J. Sloane, *The Theory of Error-Correcting Codes*. New York: Elsevier, 1977.

[13] R. W. Hamming, "Error detecting and error correcting codes," *Bell System Tech. J.*, vol. 26, no. 2, pp. 147–160, 1950.

[14] M. J. E. Golay, "Notes on digital coding," *Proc. IRE*, vol. 37, p. 657, June 1949.

[15] A. Hocquenghem, "Codes correcteurs d'erreurs," *Chiffres*, vol. 2, pp. 147–156, 1959.

[16] R. C. Bose and D. K. Ray-Chaudhuri, "On a class of error correcting binary group codes," *Info. and Control*, vol. 3, no. 1, pp. 68–79, Mar. 1960.

[17] I. S. Reed and G. Solomon, "Polynomial codes over certain finite fields," *J. SIAM*, vol. 8, no. 2, pp. 300–304, June 1960.

[18] D. E. Muller, "Application of boolean algebra to switching circuit design and to error detection," *IRE Trans. Electronic Computers*, vol. EC-3, no. 3, pp. 6–12, 1954.

[19] I. Reed, "A class of multiple-error-correcting codes and the decoding scheme," *IRE Trans. Electronic Computers*, vol. 4, no. 4, pp. 38–49, 1954.

[20] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding," in *Proc. of the IEEE Int. Conf. Commun. (ICC)*, Geneve, Switzerland, May 1993, pp. 1064–1070.

[21] R. G. Gallager, *Low-Density Parity-Check Codes.* Cambridge, MA, USA: M.I.T. Press, 1963.

[22] D. J. C. MacKay, "Good error correcting codes based on very sparse matrices," *IEEE Trans. Inform. Theory*, vol. 45, no. 2, pp. 399–431, Mar. 1999.

[23] D. A. Spielman, "Linear-time encodable and decodable error-correcting codes," *IEEE Trans. Inform. Theory*, vol. 42, no. 6, pp. 1723–1731, Nov. 1996.

[24] N. Wiberg, H.-A. Loeliger, and R. Kötter, "Codes and iterative decoding on general graphs," *European Trans. Telecomm.*, vol. 6, pp. 513–526, Sept. 1995.

[25] N. Wiberg, "Codes and decoding on general graphs," Ph.D. dissertation, Linköping University, Linköping, Sweden, 1996.

[26] R. M. Tanner, "A recursive approach to low complexity codes," *IEEE Trans. Inform. Theory*, vol. 27, no. 5, pp. 533–547, Sept. 1981.

[27] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[28] H.-A. Loeliger, "An introduction to factor graphs," *Signal Processing Magazine*, vol. 21, no. 2, pp. 28–41, Jan. 2004.

[29] M. Luby, M. Mitzenmacher, A. Shokrollahi, and D. A. Spielman, "Efficient erasure correcting codes," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 569–584, Feb. 2001.

[30] T. Richardson and R. Urbanke, "The capacity of low-density parity check codes under message-passing decoding," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 599–618, Feb. 2001.

[31] T. Richardson, A. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 619–637, Feb. 2001.

[32] S.-Y. Chung, G. D. Forney, Jr., T. Richardson, and R. Urbanke, "On the design of low-density parity-check codes within 0.0045 dB of the Shannon limit," *IEEE Commun. Lett.*, vol. 5, no. 2, pp. 58–60, Feb. 2001.

[33] A. J. Felström and K. S. Zigangirov, "Time-varying periodic convolutional codes with low-density parity-check matrix," *IEEE Trans. Inform. Theory*, vol. 45, no. 6, pp. 2181–2190, Sept. 1999.

[34] M. Lentmaier, A. Sridharan, K. S. Zigangirov, and J. D. J. Costello, "Iterative decoding threshold analysis for LDPC convolutional codes," *IEEE Trans. Inform. Theory*, vol. 56, no. 10, pp. 5274–5289, Oct. 2010.

[35] S. Kudekar, T. J. Richardson, and R. L. Urbanke, "Threshold saturation via spatial coupling: Why convolutional LDPC ensembles perform so well over the BEC," *IEEE Trans. Inform. Theory*, vol. 57, no. 2, pp. 803–834, Feb. 2011.

[36] ——, "Spatially coupled ensembles universally achieve capacity under belief propagation," *IEEE Trans. Inform. Theory*, vol. 59, no. 12, pp. 7761–7813, Dec. 2013.

[37] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inform. Theory*, vol. 55, no. 7, pp. 3051–3073, July 2009.

[38] M. El-Khamy, H.-P. Lin, and J. Lee, "Binary polar codes are optimised codes for bitwise multistage decoding," *Electronics Letters*, vol. 52, no. 13, pp. 1130–1132, June 2016.

[39] N. Stolte, "Rekursive codes mit der Plotkin-konstruktion und ihre decodierung," Ph.D. dissertation, Technische Universität Darmstadt, Darmstadt, Germany, Jan. 2002.

[40] R. J. McEliece, "Are turbo-like codes effective on nonstandard channels?" *IEEE Inform. Theory Soc. Newslett.*, vol. 51, no. 4, pp. 1–8, Dec. 2001.

[41] E. Şaşoğlu, "Polarization and polar codes," *Found. Trends Commun. Inform. Theory*, vol. 8, no. 4, pp. 259–381, 2012.

[42] S. B. Korada, "Polar codes for channel and source coding," Ph.D. dissertation, EPFL, Lausanne, Switzerland, 2009.

[43] S. H. Hassani, "Polarization and spatial coupling: Two techniques to boost performance," Ph.D. dissertation, EPFL, Lausanne, Switzerland, 2013.

[44] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge University Press, 2008.

[45] R. Mori and T. Tanaka, "Performance and construction of polar codes on symmetric binary-input memoryless channels," in *Proc. of the IEEE Int. Symposium on Inform. Theory (ISIT)*, Seoul, South Korea, July 2009, pp. 1496–1500.

[46] I. Tal and A. Vardy, "How to construct polar codes," *IEEE Trans. Inform. Theory*, vol. 59, no. 10, pp. 6562–6582, Oct. 2013.

[47] R. Pedarsani, H. Hassani, I. Tal, and E. Telatar, "On the construction of polar codes," in *Proc. of the IEEE Int. Symposium on Inform. Theory (ISIT)*, St. Petersberg, Russia, Aug. 2011, pp. 11–15.

[48] S. H. Hassani and R. Urbanke, "Universal polar codes," Dec. 2013, [Online]. Available: http://arxiv.org/pdf/1307.7223v2.pdf.

[49] E. Şaşoğlu and L. Wang, "Universal polarization," *IEEE Trans. Inform. Theory*, vol. 62, no. 6, pp. 2937–2946, June 2016.

[50] E. Arıkan, H. Kim, G. Markarian, U. Ozgur, and E. Poyraz, "Performance of short polar codes under ML decoding," in *Proc. of the ICT-Mobile Summit Conf.*, 2009.

[51] S. Kahraman and M. E. Celebi, "Code based efficient maximum-likelihood decoding of short polar codes," in *Proc. of the IEEE Int. Symposium on Inform. Theory (ISIT)*, Cambridge, MA, USA, July 2012, pp. 1967–1971.

[52] N. Goela, S. B. Korada, and M. Gastpar, "On LP decoding of polar codes," in *Proc. of the IEEE Inform. Theory Workshop (ITW)*, Cairo, Egypt, Jan. 2010, pp. 1–5.

[53] N. Hussami, S. B. Korada, and R. Urbanke, "Performance of polar codes for channel and source coding," in *Proc. of the IEEE Int. Symposium on Inform. Theory (ISIT)*, Seoul, South Korea, July 2009, pp. 1488–1492.

[54] A. Eslami and H. Pishro-Nik, "On finite-length performance of polar codes: Stopping sets, error floor, and concatenated design," *IEEE Trans. Commun.*, vol. 61, no. 3, pp. 919–929, Mar. 2013.

[55] I. Tal and A. Vardy, "List decoding of polar codes," *IEEE Trans. Inform. Theory*, vol. 61, no. 5, pp. 2213–2226, May 2015.

[56] E. Arıkan and I. E. Telatar, "On the rate of channel polarization," in *Proc. of the IEEE Int. Symposium on Inform. Theory (ISIT)*, Seoul, South Korea, July 2009, pp. 1493–1495.

[57] E. Arıkan, "A performance comparison of polar codes and Reed-Muller codes," *IEEE Commun. Lett.*, vol. 12, no. 6, pp. 447–449, June 2008.

[58] ——, "A survey of Reed-Muller codes from polar coding perspective," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Cairo, Egypt, Jan. 2010, pp. 1–5.

[59] I. Dumer, "Recursive decoding and its performance for low-rate Reed-Muller codes," *IEEE Trans. Inform. Theory*, vol. 50, no. 5, pp. 811–823, May 2004.

[60] I. Dumer and K. Shabunov, "Soft-decision decoding of Reed-Muller codes: Recursive lists," *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 1260–1266, Mar. 2006.

[61] ——, "Recursive list decoding for Reed-Muller codes and their subcodes," in *Information, Coding and Mathematics*. Springer US, 2002, vol. 687, pp. 279–298.

[62] I. Dumer, "Nested polarized codes: General design," in *14th Int. Workshop on Alg. and Comb. Coding Theory (ACCT)*, Kaliningrad, Russia, Sept. 2014, pp. 139–144.

[63] ——, "Nested polarized codes: Decoding and node selection," in *14th Int. Workshop on Alg. and Comb. Coding Theory (ACCT)*, Kaliningrad, Russia, Sept. 2014, pp. 145–150.

[64] B. Li, H. Shen, and D. Tse, "A RM-polar codes," July 2014, [Online]. Available: `http://arxiv.org/abs/1407.5483`.

[65] I. Dumer and P. G. Farrell, "Erasure correction performance of linear block codes," in *Algebraic Coding*. Springer, 1994, pp. 316–326.

[66] C. Carlet and P. Gaborit, "On the construction of balanced boolean functions with a good algebraic immunity," in *Proc. of the IEEE Int. Symposium on Inform. Theory (ISIT)*, Sept. 2005, pp. 1101–1105.

[67] F. Didier, "A new upper bound on the block error probability after decoding over the erasure channel," *IEEE Trans. Inform. Theory*, vol. 52, no. 10, pp. 4496–4503, Oct. 2006.

[68] E. Abbe, A. Shpilka, and A. Wigderson, "Reed-Muller codes for random erasures and errors," *IEEE Trans. Inform. Theory*, vol. 61, no. 10, pp. 5229–5252, Oct. 2015.

[69] ——, "Reed-Muller codes for random erasures and errors," in *Proc. of the Annual ACM Symposium on Theory of Computing (STOC)*, Portland, OR, USA, June 2015, pp. 297–306.

[70] A. Barg and G. D. Forney, Jr., "Random codes: Minimum distances and error exponents," *IEEE Trans. Inform. Theory*, vol. 48, pp. 2568–2573, Sep. 2002.

[71] J.-P. Tillich and G. Zémor, "Discrete isoperimetric inequalities and the probability of a decoding error," *Combinatorics, Probability and Computing*, vol. 9, no. 5, pp. 465–479, Sept. 2000.

[72] A. Montanari, "Finite size scaling and metastable states of good codes," in *Proc. of the Allerton Conf. on Commun., Control, and Computing*, Monticello, IL, USA, Oct. 2001.

[73] R. L. Dobrushin, "Mathematical problems in the Shannon theory of optimal coding of information," in *Proc. 4th Berkeley Symp. Mathematics, Statistics, and Probability*, vol. 1, 1961, pp. 211–252.

[74] V. Strassen, "Asymptotische abschätzungen in Shannon's informationstheorie," in *Trans. 3rd Prague Conf. Inf. Theory*, 1962, pp. 689–723.

[75] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *IEEE Trans. Inform. Theory*, vol. 55, no. 11, pp. 4947–4966, Nov. 2009.

[76] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite block-length regime," *IEEE Trans. Inform. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[77] A. Amraoui, A. Montanari, T. Richardson, and R. Urbanke, "Finite-length scaling for iteratively decoded LDPC ensembles," *IEEE Trans. Inform. Theory*, vol. 55, no. 2, pp. 473–498, Feb. 2009.

[78] Y. Altug and A. B. Wagner, "Moderate deviations in channel coding," *IEEE Trans. Inform. Theory*, vol. 60, no. 8, pp. 4417–4426, Aug. 2014.

[79] E. Şaşoğlu, I. E. Telatar, and E. Arıkan, "Polarization for arbitrary discrete memoryless channels," in *Proc. of the IEEE Inform. Theory Workshop (ITW)*, Taormina, Italy, Oct. 2009, pp. 144–148.

[80] R. Mori and T. Tanaka, "Channel polarization on q-ary discrete memoryless channels by arbitrary kernel," in *Proc. of the IEEE Int. Symposium on Inform. Theory (ISIT)*, Austin, TX, USA, June 2010, pp. 894–898.

[81] I. Tal, A. Sharov, and A. Vardy, "Constructing polar codes for non-binary alphabets and MACs," in *Proc. of the IEEE Int. Symposium on Inform. Theory (ISIT)*, Cambridge, MA, USA, July 2012, pp. 2142–2146.

[82] W. Park and A. Barg, "Polar codes for q-ary channels, $q = 2^r$," *IEEE Trans. Inform. Theory*, vol. 59, no. 2, pp. 955–969, Feb. 2013.

[83] A. G. Sahebi and S. S. Pradhan, "Multilevel channel polarization for arbitrary discrete memoryless channels," *IEEE Trans. Inform. Theory*, vol. 59, no. 12, pp. 7839–7857, Dec. 2013.

[84] R. Nasser and I. E. Telatar, "Polar codes for arbitrary DMCs and arbitrary MACs," *IEEE Trans. Inform. Theory*, vol. 62, no. 6, pp. 2917–2936, June 2016.

[85] R. Nasser, "Ergodic theory meets polarization I: A foundation of polarization theory," in *Proc. of the IEEE Int. Symposium on Inform. Theory (ISIT)*, Hong Kong, June 2015, pp. 2451–2455.

[86] E. Abbe and A. Barron, "Polar coding schemes for the AWGN channel," in *Proc. of the IEEE Int. Symposium on Inform. Theory (ISIT)*, St. Petersberg, Russia, Aug. 2011, pp. 194–198.

[87] H. Uchikawa, K. Kasai, and K. Sakaniwa, "Design and performance of rate-compatible non-binary LDPC convolutional codes," *IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences*, pp. 2135–2143, 2011.

[88] A. Piemontese, A. G. Amat, and G. Colavolpe, "Nonbinary spatially-coupled LDPC codes on the binary erasure channel," in *Proc. of the IEEE Int. Conf. Commun. (ICC)*, Budapest, Hungary, June 2013, pp. 3270–3274.

[89] I. Andriyanova and A. G. Amat, "Threshold saturation for nonbinary SC-LDPC codes on the binary erasure channel," *IEEE Trans. Inform. Theory*, vol. 62, no. 5, pp. 2622–2638, May 2016.

[90] L. Wei, T. Koike-Akino, D. G. M. Mitchell, T. E. Fuja, and D. J. Costello Jr, "Threshold analysis of non-binary spatially-coupled LDPC codes with windowed decoding," Mar. 2014, [Online]. Available: `http://arxiv.org/abs/1403.3583`.

[91] E. Arıkan, "Source polarization," in *Proc. of the IEEE Int. Symposium on Inform. Theory (ISIT)*, Austin, TX, USA, June 2010, pp. 899–903.

[92] H. S. Cronie and S. B. Korada, "Lossless source coding with polar codes," in *Proc. of the IEEE Int. Symposium on Inform. Theory (ISIT)*, Austin, USA, June 2010, pp. 904–908.

[93] S. B. Korada and R. Urbanke, "Polar codes are optimal for lossy source coding," *IEEE Trans. Inform. Theory*, vol. 56, no. 4, pp. 1751–1768, Apr. 2010.

[94] E. Arikan, "Polar coding for the Slepian-Wolf problem based on monotone chain rules," in *Proc. of the IEEE Int. Symposium on Inform. Theory (ISIT)*, Cambridge, MA, USA, July 2012, pp. 571–575.

[95] L. Wang and Y.-H. Kim, "Linear code duality between channel coding and Slepian-Wolf coding," in *Proc. of the Allerton Conf. on Commun., Control, and Computing*, Monticello, IL, USA, Oct. 2015, pp. 147–152.

[96] E. Şaşoğlu, I. E. Telatar, and E. Yeh, "Polar codes for the two-user binary-input multiple-access channel," in *Proc. of the IEEE Inform. Theory Workshop (ITW)*, Cairo, Egypt, Jan. 2010, pp. 1–5.

[97] E. Abbe and I. E. Telatar, "Polar codes for the m-user multiple access channel," *IEEE Trans. Inform. Theory*, vol. 58, no. 8, pp. 5437–5448, Aug. 2012.

[98] H. Mahdavifar, M. El-Khamy, J. Lee, and I. Kang, "Achieving the uniform rate region of general multiple access channels by polar coding," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 467–478, Feb. 2016.

[99] M. Andersson, R. F. Schaefer, T. J. Oechtering, and M. Skoglund, "Polar coding for bidirectional broadcast channels with common and confidential messages," *IEEE J. Select. Areas Commun.*, vol. 31, no. 9, pp. 1901–1908, Sept. 2013.

[100] N. Goela, E. Abbe, and M. Gastpar, "Polar codes for broadcast channels," *IEEE Trans. Inform. Theory*, vol. 61, no. 2, pp. 758–782, Feb. 2015.

[101] K. Appaiah, O. Koyluoglu, and S. Vishwanath, "Polar alignment for interference networks," in *Proc. of the Allerton Conf. on Commun., Control, and Computing*, Monticello, IL, USA, Sept. 2011, pp. 240–246.

[102] L. Wang and E. Şaşoğlu, "Polar coding for interference networks," in *Proc. of the IEEE Int. Symposium on Inform. Theory (ISIT)*, Honolulu, HI, USA, July 2014, pp. 311–315.

[103] M. Andersson, V. Rathi, R. Thobaben, J. Kliewer, and M. Skoglund, "Nested polar codes for wiretap and relay channels," *IEEE Comm. Lett.*, vol. 14, no. 8, pp. 752–754, Aug. 2010.

[104] M. Karzand, "Polar codes for degraded relay channels," in *Proc. Intern. Zurich Seminar on Comm.*, Zurich, Switzerland, Feb. 2012, pp. 59–62.

[105] R. Blasco-Serrano, R. Thobaben, M. Andersson, V. Rathi, and M. Skoglund, "Polar codes for cooperative relaying," *IEEE Trans. Commun.*, vol. 60, no. 11, pp. 3263–3273, Nov. 2012.

[106] L. Wang, "Polar coding for relay channels," in *Proc. of the IEEE Int. Symposium on Inform. Theory (ISIT)*, Hong Kong, June 2015, pp. 1532–1536.

[107] H. Mahdavifar and A. Vardy, "Achieving the secrecy capacity of wiretap channels using polar codes," *IEEE Trans. Inform. Theory*, vol. 57, no. 10, pp. 6428–6443, Oct. 2011.

[108] O. Koyluoglu and H. E. Gamal, "Polar coding for secure transmission and key agreement," *IEEE Trans. Inform. Forensics Security*, vol. 7, no. 5, pp. 1472–1483, Oct. 2012.

[109] E. Hof and S. Shamai, "Secrecy-achieving polar-coding," in *Proc. of the IEEE Inform. Theory Workshop (ITW)*, Dublin, Ireland, September 2010, pp. 1–5.

[110] E. Şaşoğlu and A. Vardy, "A new polar coding scheme for strong security on wiretap channels," in *Proc. of the IEEE Int. Symposium on Inform. Theory (ISIT)*, Istanbul, Turkey, July 2013, pp. 1117–1121.

[111] D. Burshtein and A. Strugatski, "Polar write once memory codes," *IEEE Trans. Inform. Theory*, vol. 59, no. 8, pp. 5088–5101, August 2013.

[112] E. Hof, I. Sason, S. Shamai, and C. Tian, "Capacity-achieving polar codes for arbitrarily-permuted parallel channels," *IEEE Trans. Inform. Theory*, vol. 59, no. 3, pp. 1505–1516, Mar. 2013.

[113] A. G. Sahebi and S. S. Pradhan, "Polar codes for some multi-terminal communications problems," in *Proc. of the IEEE Int. Symposium on Inform. Theory (ISIT)*, Honolulu, HI, USA, July 2014, pp. 316–320.

[114] Q. Shi, L. Song, C. Tian, J. Chen, and S. Dumitrescu, "Polar codes for multiple descriptions," *IEEE Trans. Inform. Theory*, vol. 61, no. 1, pp. 107–119, Jan. 2015.

[115] L. Wang, "Channel coding techniques for network communication," Ph.D. dissertation, UCSD, San Diego, California, USA, 2015.

[116] E. E. Gad, Y. Li, J. Kliewer, M. Langberg, A. Jiang, and J. Bruck, "Asymmetric error correction and flash-memory rewriting using polar codes," *IEEE Trans. Inform. Theory*, vol. 62, no. 7, pp. 4024–4038, July 2016.

[117] Y.-P. Wei and S. Ulukus, "Polar coding for the general wiretap channel with extensions to multiuser scenarios," *IEEE J. Select. Areas Commun.*, vol. 34, no. 2, pp. 278–291, Feb. 2016.

[118] T. C. Gulcu and A. Barg, "Achieving secrecy capacity of the wiretap channel and broadcast channel with a confidential component," in *Proc. of the IEEE Inform. Theory Workshop (ITW)*, Jerusalem, Israel, May 2015, pp. 1–5.

[119] R. A. Chou and M. R. Bloch, "Polar coding for the broadcast channel with confidential messages: A random binning analogy," *IEEE Trans. Inform. Theory*, vol. 62, no. 5, pp. 2410–2429, May 2016.

[120] R. G. Gallager, *Information Theory and Reliable Communication.* New York: Wiley, 1968.

[121] J. Honda and H. Yamamoto, "Polar coding without alphabet extension for asymmetric models," *IEEE Trans. Inform. Theory*, vol. 59, no. 12, pp. 7829–7838, Dec. 2013.

[122] G. Böcherer and R. Mathar, "Operating LDPC codes with zero shaping gap," in *Proc. of the IEEE Inform. Theory Workshop (ITW)*, Paraty, Brazil, Oct. 2011, pp. 330–334.

[123] M. Mondelli, S. H. Hassani, and R. Urbanke, "Unified scaling of polar codes: Error exponent, scaling exponent, moderate deviations, and error floors," in *Proc. of the IEEE Int. Symposium on Inform. Theory (ISIT)*, Hong Kong, June 2015, pp. 1422–1426.

[124] ——, "Unified scaling of polar codes: Error exponent, scaling exponent, moderate deviations, and error floors," accepted to *IEEE Trans. Inform. Theory*, Oct. 2016.

[125] S. H. Hassani, R. Mori, T. Tanaka, and R. Urbanke, "Rate-dependent analysis of the asymptotic behavior of channel polarization," *IEEE Trans. Inform. Theory*, vol. 59, no. 4, pp. 2267–2276, Apr. 2013.

[126] S. B. Korada, A. Montanari, I. E. Telatar, and R. Urbanke, "An empirical scaling law for polar codes," in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*, Austin, TX, USA, June 2010, pp. 884–888.

[127] V. Guruswami and P. Xia, "Polar codes: Speed of polarization and polynomial gap to capacity," *IEEE Trans. Inform. Theory*, vol. 61, no. 1, pp. 3–16, Jan. 2015.

[128] S. H. Hassani, K. Alishahi, and R. Urbanke, "Finite-length scaling for polar codes," *IEEE Trans. Inform. Theory*, vol. 60, no. 10, pp. 5875–5898, Oct. 2014.

[129] D. Goldin and D. Burshtein, "Improved bounds on the finite length scaling of polar codes," *IEEE Trans. Inform. Theory*, vol. 60, no. 11, pp. 6966–6978, Nov. 2014.

[130] M. Mondelli, S. H. Hassani, and R. Urbanke, "Scaling exponent of list decoders with applications to polar codes," in *Proc. of the IEEE Inform. Theory Workshop (ITW)*, Sevilla, Spain, Sept. 2013, pp. 1–5.

[131] ——, "Scaling exponent of list decoders with applications to polar codes," *IEEE Trans. Inform. Theory*, vol. 61, no. 9, pp. 4838–4851, Sept. 2015.

[132] P. Elias, "List decoding for noisy channels," Institute of Radio Engineers (now IEEE), Tech. Rep., 1957.

[133] J. M. Wozencraft, "List decoding," Research Laboratory of Electronics, Massachusetts Institute of Technology, Tech. Rep., 1958.

[134] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. I," *Inform. Contr.*, vol. 10, no. 1, pp. 65–103, Jan. 1967.

[135] G. D. J. Forney, "Exponential error bounds for erasure, list, and decision feedback schemes," *IEEE Trans. Inform. Theory*, vol. 14, no. 2, pp. 206–220, Mar. 1968.

[136] E. Hof, I. Sason, and S. Shamai, "Performance bounds for erasure, list, and decision feedback schemes with linear block codes," *IEEE Trans. Inform. Theory*, vol. 56, no. 8, pp. 3754–3778, Aug. 2010.

[137] V. F. Kolchin, *Random Graphs*. Cambridge University Press, 1999.

[138] C. M. Fortuin, P. W. Kasteleyn, and J. Ginibre, "Correlation inequalities on some partially ordered sets," *Commun. Math. Phys.*, vol. 22, pp. 89–103, 1971.

[139] S. B. Korada and R. Urbanke, "Exchange of limits: Why iterative decoding works," *IEEE Trans. Inform. Theory*, vol. 57, no. 4, pp. 2169–2187, Apr. 2011.

[140] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Ann. Math. Statist.*, vol. 23, no. 4, pp. 493–507, Dec. 1952.

[141] N. Alon, J. Spencer, and P. Erdös, *The Probabilistic Method*. John Wiley & Sons, Inc., 2000.

[142] C. J. Preston, "A generalization of the FKG inequalities," *Commun. Math. Phys.*, vol. 36, pp. 233–241, 1974.

[143] M. B. Parizi and I. E. Telatar, "On correlation between polarized BECs," in *Proc. of the IEEE Int. Symposium on Inform. Theory (ISIT)*, Istanbul, Turkey, July 2013, pp. 784–788.

[144] L. Comtet, *Advanced Combinatorics: The Art of Finite and Infinite Expansions*. D. Reidel, Dordrecht, 1974.

[145] M. Mondelli, S. H. Hassani, I. Sason, and R. Urbanke, "Achieving Marton's region for broadcast channels using polar codes," in *Proc. of the IEEE Int. Symposium on Inform. Theory (ISIT)*, Honolulu, HI, USA, July 2014, pp. 306–310.

[146] ——, "Achieving Marton's region for broadcast channels using polar codes," *IEEE Trans. Inform. Theory*, vol. 61, no. 2, pp. 783–800, Feb. 2015.

[147] P. P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Inform. Theory*, vol. 19, no. 2, pp. 197–207, March 1973.

[148] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inform. Theory*, vol. 25, no. 3, pp. 306–311, May 1979.

[149] T. M. Cover, "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. 18, no. 1, pp. 2–14, Jan. 1972.

[150] L. Wang, E. Şaşoğlu, B. Bandemer, and Y.-H. Kim, "A comparison of superposition coding schemes," in *Proc. of the IEEE Int. Symposium on Inform. Theory (ISIT)*, Istanbul, Turkey, July 2013, pp. 2970–2974.

[151] S. I. Gelfand and M. S. Pinsker, "Capacity of a broadcast channel with one deterministic component," *Problemy Peredachi Informatsii*, vol. 16, no. 1, pp. 24–34, 1980.

[152] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

[153] G. Kramer, "Topics in multi-user information theory," *Found. Trends Commun. Inf. Theory*, vol. 4, no. 4-5, pp. 265–444, Apr. 2007.

[154] S. H. Hassani, S. B. Korada, and R. Urbanke, "The compound capacity of polar codes," in *Proc. of the Allerton Conf. on Commun., Control, and Computing*, Monticello, IL, USA, 2009, pp. 16–21.

[155] Y. Liang, "Multiuser communications with relaying and user cooperation," Ph.D. dissertation, UIUC, Urbana-Champaign, Illinois, USA, 2005.

[156] Y. Liang and G. Kramer, "Rate regions for relay broadcast channels," *IEEE Trans. Inform. Theory*, vol. 53, no. 10, pp. 3517–3535, Oct. 2007.

[157] Y. Liang, G. Kramer, and H. V. Poor, "On the equivalence of two achievable regions for the broadcast channel," *IEEE Trans. Inform. Theory*, vol. 57, no. 1, pp. 95–100, Jan. 2011.

[158] A. A. Gohari and V. Anantharam, "Evaluation of Marton's inner bound for the general broadcast channel," *IEEE Trans. Inform. Theory*, vol. 58, no. 2, pp. 608–619, Feb. 2012.

[159] Y. Geng, V. Jog, C. Nair, and Z. V. Wang, "An information inequality and evaluation of Marton's inner bound for binary-input broadcast channels," *IEEE Trans. Inform. Theory*, vol. 59, no. 7, pp. 4095–4105, July 2013.

[160] A. Gohari, C. Nair, and V. Anantharam, "Improved cardinality bounds on the auxiliary random variables in Marton's inner bound," in *Proc. of the IEEE Int. Symposium on Inform. Theory (ISIT)*, Istanbul, Turkey, July 2013, pp. 1272–1276.

[161] Y. Geng, A. Gohari, C. Nair, and Y. Yu, "On Marton's inner bound and its optimality for classes of product broadcast channels," *IEEE Trans. Inform. Theory*, vol. 60, no. 1, pp. 22–41, Jan. 2014.

[162] Y. Geng and C. Nair, "The capacity region of the two-receiver Gaussian vector broadcast channel with private and common messages," *IEEE Trans. Inform. Theory*, vol. 60, no. 4, pp. 2087–2104, Apr. 2014.

[163] M. Salehi, "Cardinality bounds on auxiliary variables in multiple-user theory via the method of Ahlswede and Korner," Stanford University, Tech. Rep. 33, 1978.

[164] Y. Geng, C. Nair, S. Shamai, and Z. V. Wang, "On broadcast channels with binary inputs and symmetric outputs," *IEEE Trans. Inform. Theory*, vol. 59, no. 11, pp. 6980–6989, Nov. 2013.

[165] D. Sutter, J. M. Renes, F. Dupuis, and R. Renner, "Achieving the capacity of any DMC using only polar codes," in *Proc. of the IEEE Inform. Theory Workshop (ITW)*, Lausanne, Switzerland, Sept. 2012, pp. 114–118.

[166] M. Mondelli, S. H. Hassani, and R. Urbanke, "How to achieve the capacity of asymmetric channels," in *Proc. of the Allerton Conf. on Commun., Control, and Computing*, Monticello, IL, USA, Oct. 2014, pp. 789–796.

[167] ——, "How to achieve the capacity of asymmetric channels," submitted to *IEEE Trans. Inform. Theory*, Sept. 2014. [Online]. Available: `http://arxiv.org/abs/1406.7373`.

[168] E. E. Majani and H. Rumsey, Jr., "Two results on binary-input discrete memoryless channels," in *Proc. of the IEEE Int. Symposium on Inform. Theory (ISIT)*, Budapest, Hungary, June 1991, p. 104.

[169] N. Shulman and M. Feder, "The uniform distribution as a universal prior," *IEEE Trans. Inform. Theory*, vol. 50, no. 6, pp. 1356–1362, June 2004.

[170] X.-B. Liang, "On a conjecture of Majani and Rumsey," in *Proc. of the IEEE Int. Symposium on Inform. Theory (ISIT)*, Chicago, USA, June 2004, p. 62.

[171] G. Caire, S. Shamai, and S. Verdú, "Noiseless data compression with low-density parity-check codes," *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, vol. 66, pp. 263–284, 2004.

[172] J. Hou, P. H. Siegel, L. B. Milstein, and H. D. Pfister, "Capacity-approaching bandwidth-efficient coded modulation schemes based on low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 49, no. 9, pp. 2141–2155, Sept. 2003.

[173] C.-C. Wang, S. R. Kulkarni, and H. V. Poor, "Density evolution for asymmetric memoryless channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4216–4236, Dec. 2005.

[174] L. Wang and Y.-H. Kim, "Linear code duality between channel coding and Slepian-Wolf coding," in *Proc. of the Allerton Conf. on Commun., Control, and Computing*, Monticello, IL, USA, Oct. 2015, pp. 147–152.

[175] K. Kasai and K. Sakaniwa, "Spatially-coupled MacKay-Neal codes and Hsu-Anastasopoulos codes," *IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 94, no. 11, pp. 2161–2168, Nov. 2011.

[176] D. G. M. Mitchell, K. Kasai, M. Lentmaier, and D. J. Costello, Jr., "Asymptotic analysis of spatially coupled MacKay-Neal and Hsu-Anastasopoulos LDPC codes," in *Proc. of the IEEE Int. Symposium on Inform. Theory and its Applications (ISITA)*, Honolulu, HI, USA, Oct. 2012, pp. 337–341.

[177] J. B. Soriaga and P. H. Siegel, "On distribution shaping codes for partial-response channels," in *Proc. of the Allerton Conf. on Commun., Control, and Computing*, Monticello, IL, USA, 2003.

[178] S. Ciliberti and M. Mézard, "The theoretical capacity of the parity source coder," *Journal of Statistical Mechanics: Theory and Experiment*, no. 10, Oct. 2005.

[179] V. Aref, N. Macris, and M. Vuffray, "Approaching the rate-distortion limit with spatial coupling, belief propagation, and decimation," *IEEE Trans. Inform. Theory*, vol. 61, no. 7, pp. 3954–3979, July 2015.

[180] S. Kumar, A. Vem, K. Narayanan, and H. D. Pfister, "Spatially-coupled codes for side-information problems," in *Proc. of the IEEE Int. Symposium on Inform. Theory (ISIT)*, Honolulu, HI, USA, July 2014, pp. 516–520.

[181] J. Muramatsu and S. Miyake, "Hash property and coding theorems for sparse matrices and maximum-likelihood coding," *IEEE Trans. Inform. Theory*, vol. 56, no. 5, pp. 2143–2167, May 2010.

[182] J. Muramatsu, "Channel coding and lossy source coding using a generator of constrained random numbers," *IEEE Trans. Inform. Theory*, vol. 60, no. 5, pp. 2667–2686, May 2014.

[183] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York: Wiley, 2006.

[184] M. Cheraghchi, "Applications of derandomization in coding theory," Ph.D. dissertation, EPFL, Lausanne, Switzerland, July 2010.

[185] E. Abbe, "Randomness and dependencies extraction via polarization, with applications to slepian-wolf coding and secrecy," *IEEE Trans. Inform. Theory*, vol. 61, no. 5, pp. 2388–2398, May 2015.

[186] Z. Zhang, "Estimating mutual information via Kolmogorov distance," *IEEE Trans. Inform. Theory*, vol. 53, no. 9, pp. 3280–3282, Sept. 2007.

[187] I. Csiszár and J. Körner, *Information theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2011.

[188] M. Mondelli, S. H. Hassani, and R. Urbanke, "From polar to Reed-Muller codes: A technique to improve the finite-length performance," in *Proc. of the IEEE Int. Symposium on Inform. Theory (ISIT)*, Honolulu, HI, USA, July 2014, pp. 131–135.

[189] ——, "From polar to Reed-Muller codes: A technique to improve the finite-length performance," *IEEE Trans. Commun.*, vol. 62, no. 9, pp. 3084–3091, Sept. 2014.

[190] S. Kudekar, M. Mondelli, E. Şaşoğlu, and R. Urbanke, "Reed-Muller codes achieve capacity on the binary erasure channel under MAP decoding," May 2015. [Online]. Available: `http://arxiv.org/abs/1505.05831`.

[191] S. Kumar and H. D. Pfister, "Reed-Muller codes achieve capacity on erasure channels," May 2015. [Online]. Available: `http://arxiv.org/abs/1505.05123`.

[192] S. Kudekar, S. Kumar, M. Mondelli, H. D. Pfister, E. Şaşoğlu, and R. Urbanke, "Reed-Muller codes achieve capacity on erasure channels," submitted to *IEEE Trans. Inform. Theory*, Jan. 2016. [Online]. Available: `http://arxiv.org/abs/1601.04689`.

[193] ——, "Reed-Muller codes achieve capacity on erasure channels," in *Proc. of the Annual ACM Symposium on Theory of Computing (STOC)*, Boston, MA, USA, June 2016, pp. 658–669.

[194] S. Kumar, "Capacity-achieving coding mechanisms: Spatial coupling and group symmetry," Ph.D. dissertation, Texas A&M University, College Station, Texas, USA, 2015.

[195] R. Ahlswede and G. Dueck, "Good codes can be produced by a few permutations," *IEEE Trans. Inform. Theory*, vol. 28, no. 3, pp. 430–443, May 1982.

[196] J. T. Coffey and R. M. Goodman, "Any code of which we cannot think is good," *IEEE Trans. Inform. Theory*, vol. 36, no. 6, pp. 1453–1461, Nov. 1990.

[197] T. Kasami, S. Lin, and W. W. Peterson, "New generalizations of the Reed-Muller codes–I: Primitive codes," *IEEE Trans. Inform. Theory*, vol. 14, no. 2, pp. 189–199, Mar. 1968.

[198] G. A. Margulis, "Probabilistic characteristics of graphs with large connectivity," *Problems of Inform. Transm.*, vol. 10, no. 2, pp. 101–108, 1974.

[199] L. Russo, "An approximate zero-one law," *Prob. Th. and Related Fields*, vol. 61, no. 1, pp. 129–139, 1982.

[200] M. Talagrand, "Isoperimetry, logarithmic sobolev inequalities on the discrete cube, and margulis' graph connectivity theorem," *Geometric & Functional Analysis*, vol. 3, no. 3, pp. 295–314, 1993.

[201] ——, "On Russo's approximate zero-one law," *Ann. Probab.*, vol. 22, no. 3, pp. 1576–1587, 1994.

[202] E. Friedgut and G. Kalai, "Every monotone graph property has a sharp threshold," *Proc. Amer. Math. Soc.*, vol. 124, no. 10, pp. 2993–3002, 1996.

[203] E. Friedgut and J. Bourgain, "Sharp thresholds of graph properties, and the k-sat problem," *J. Amer. Math. Soc.*, vol. 12, no. 4, pp. 1017–1054, 1999.

[204] I. Dinur and S. Safra, "On the hardness of approximating minimum vertex cover," *Ann. Math.*, vol. 162, no. 1, pp. 439–485, July 2005.

[205] G. Zémor, "Threshold effects in codes," in *Algebraic Coding*. Springer, 1994, pp. 278–286.

[206] J.-P. Tillich and G. Zemor, "The Gaussian isoperimetric inequality and decoding error probabilities for the Gaussian channel," *IEEE Trans. Inform. Theory*, vol. 50, no. 2, pp. 328–331, Feb. 2004.

[207] G. Kalai and S. Safra, "Threshold phenomena and influence with some perspectives from mathematics, computer science, and economics," *Comp. Complexity and Stat. Phy., Santa Fe Institute Studies in Sci. of Complexity*, 2005.

[208] S. Boucheron, G. Lugosi, and P. Massart, *Concentration inequalities: A nonasymptotic theory of independence*. Oxford University Press, 2013.

[209] D. Achlioptas, A. Naor, and Y. Peres, "Rigorous location of phase transitions in hard optimization problems," *Nature*, vol. 435, pp. 759–764, June 2005.

[210] A. Coja-Oghlan, "The asymptotic k-SAT threshold," in *Proc. of the Annual ACM Symposium on Theory of Computing (STOC)*, June 2014, pp. 804–813.

[211] J. Ding, A. Sly, and N. Sun, "Proof of the satisfiability conjecture for large k," in *Proc. of the Annual ACM Symposium on Theory of Computing (STOC)*, June 2015, pp. 59–68.

[212] S. ten Brink, "Convergence of iterative decoding," *Electron. Lett.*, vol. 35, no. 10, pp. 806–808, May 1999.

[213] A. Ashikhmin, G. Kramer, and S. ten Brink, "Extrinsic information transfer functions: Model and erasure channel property," *IEEE Trans. Inform. Theory*, vol. 50, no. 11, pp. 2657–2673, Nov. 2004.

[214] C. Méasson, A. Montanari, and R. Urbanke, "Maxwell construction: The hidden bridge between iterative and maximum a posteriori decoding," *IEEE Trans. Inform. Theory*, vol. 54, no. 12, pp. 5277–5307, Dec. 2008.

[215] S. Kudekar, S. Kumar, M. Mondelli, H. D. Pfister, and R. Urbanke, "Comparing the bit-MAP and block-MAP decoding thresholds of Reed-Muller codes on BMS channels," accepted at *IEEE Int. Symposium on Inform. Theory (ISIT)*, Barcelona, Spain, July 2016.

[216] R. Saptharishi, A. Shpilka, and B. L. Volk, "Efficiently decoding Reed-Muller codes from random errors," in *Proc. of the Annual ACM Symposium on Theory of Computing (STOC)*, Boston, MA, USA, June 2016, pp. 227–235.

[217] R. B. Ash, *Information theory*. Dover Publications, 1990.

[218] J. Bourgain and G. Kalai, "Influences of variables and threshold intervals under group symmetries," *Geometric & Functional Analysis*, vol. 7, no. 3, pp. 438–461, 1997.

[219] N. J. A. Sloane and E. Berlekamp, "Weight enumerator for second-order Reed-Muller codes," *IEEE Trans. Inform. Theory*, vol. 16, no. 6, pp. 745–751, Nov. 1970.

[220] T. Kasami and N. Tokura, "On the weight structure of Reed-Muller codes," *IEEE Trans. Inform. Theory*, vol. 16, no. 6, pp. 752–759, Nov. 1970.

[221] T. Kasami, N. Tokura, and S. Azumi, "On the weight enumeration of weights less than 2.5d of Reed-Muller codes," *Inform. and Control*, vol. 30, no. 4, pp. 380–395, 1976.

[222] T. Kaufman, S. Lovett, and E. Porat, "Weight distribution and list-decoding size of Reed-Muller codes," *IEEE Trans. Inform. Theory*, vol. 58, no. 5, pp. 2689–2696, May 2012.

[223] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

[224] S. L. Fong and V. Y. F. Tan, "On the scaling exponent of polar codes for binary-input energy-harvesting channels," Apr. 2016, [Online]. Available: `http://arxiv.org/abs/1601.01089`.

[225] H. D. Pfister and R. Urbanke, "Near-optimal finite-length scaling for polar codes over large alphabets," May 2016. [Online]. Available: `http://arxiv.org/abs/1605.01997`.

[226] S. B. Korada, E. Şaşoğlu, and R. Urbanke, "Polar codes: Characterization of exponent, bounds, and constructions," *IEEE Trans. Inform. Theory*, vol. 56, no. 12, pp. 6253–6264, Dec. 2010.

[227] A. Fazeli and A. Vardy, "On the scaling exponent of binary polarization kernels," in *Proc. of the Allerton Conf. on Commun., Control, and Computing*, Monticello, IL, USA, Oct. 2014, pp. 797–804.

[228] S. Kumar, C. R., and H. D. Pfister, "Beyond double transitivity: Capacity-achieving cyclic codes on erasure channels," accepted at *IEEE Inform. Theory Workshop (ITW)*, Cambridge, UK, Sept. 2016.

[229] ——, "Reed-Muller codes achieve capacity on the quantum erasure channel," accepted at *IEEE Int. Symposium on Inform. Theory (ISIT)*, Barcelona, Spain, July 2016.

Curriculum Vitae

Marco Mondelli

Communication Theory Laboratory
École Polytechnique Fédérale de Lausanne
Station 14, 1015 Lausanne
Switzerland

+41 21 69 37514
marco.mondelli@epfl.ch
http://people.epfl.ch/marco.mondelli

Research Interests

- Coding theory
- Machine learning
- Information theory
- Wireless communication systems

Education

Sept. 2012 - Nov. 2016 **Ph.D. in Computer and Communication Sciences**
École Polytechnique Fédérale de Lausanne, Switzerland
Dissertation title: *From Polar to Reed-Muller Codes: Unified Scaling, Non-standard Channels, and a Proven Conjecture*
Advisor: *Prof. Rüdiger Urbanke.*

Oct. 2010 - July 2013 **Honors College Master's Student in Engineering**
Sant'Anna School of Advanced Studies, Italy
Grade: *100/100 cum laude.*

Oct. 2010 - July 2012 **Master's Degree in Telecommunications Engineering**
University of Pisa, Italy
Grade: *110/110 cum laude.*

Oct. 2007 - Nov. 2010 **Honors College Bachelor's Student in Engineering**
Sant'Anna School of Advanced Studies, Italy
Grade: *100/100 cum laude.*

Oct. 2007 - July 2010 **Bachelor's Degree in Telecommunications Engineering**
University of Pisa, Italy
Grade: *110/110 cum laude.*

Honors and Awards

2016 **Early Postdoc.Mobility Fellowship**, Swiss National Science Foundation.

2016 **STOC Best Paper Award** for "Reed-Muller Codes Achieve Capacity on Erasure Channels" (with S. Kudekar, S. Kumar, H. D. Pfister, E. Şaşoğlu, and R. Urbanke).

2016 **2nd Place** in the **Shannon Centennial Student Competition**.

2015 **IEEE Jack Keil Wolf ISIT Student Paper Award** for "Unified Scaling of Polar Codes: Error Exponent, Scaling Exponent, Moderate Deviations, and Error Floors".

2015 **Dan David Prize Scholarship**.

2014 **Master Thesis Award "Matteo Carmassi" for Innovation**.

2014 **I&C Outstanding Teaching Assistant Award**, EPFL.

2012 **Departmental Fellowship**, EPFL.

Research Experience

Aug. - Dec. 2015 **Visiting Graduate Student**
Information Systems Laboratory, Stanford University, USA
Advisor: *Prof. Andrea Montanari*.

Mar. - Apr. 2015 **Visiting Graduate Student**
Simons Institute for the Theory of Computing, Berkeley, USA
Program: *Information Theory*.

Aug. - Nov. 2011 **Intern**
Center for Signal and Image Processing, Georgia Institute of Technology, USA
Advisor: *Prof. Xiaoli Ma*.

Mar. - Apr. 2010 **Intern**
Centre de Mathématique et de Leurs Applications, École Normale Supérieure de Cachan, France
Advisor: *Prof. Jean-Michel Morel*.

Teaching Experience

Teaching Assistant at EPFL

- Random Walks, Spring 2015 and Spring 2016.
- Discrete Structures, Fall 2013 and Fall 2014.
- Graph Theory Applications, Spring 2014.
- Probability and Statistics, Spring 2013.

Student Project Co-Supervisor at EPFL

- Stefano Olivotto, "Feedback schemes to improve the finite-length performance of polar codes", Master thesis, Spring 2016.
- Nadim Ghaddar, "Extremes of information combining – Characterization of the BSC", Master semester project, Spring 2015.
- Georg Schölly, "Explorations on a new upper bound on the capacity for the primitive relay channel", Master semester project, Spring 2015.
- Fangyu Ye, "Extremes of information combining – Solution for the BEC", Master semester project, Spring 2015.
- Kareem Attiah, "Demos for LDPC codes and density evolution library", internship, Summer 2014.
- Frédéric Sabatier, "Polar coding tutorial", internship, Summer 2014.

Professional Service

Event Organization

- *Summer School on Information Processing for Large Networks (IPLN)*, Les Diablerets, Switzerland, June 2015.

Journal Review

- IEEE Communications Letters
- IEEE Journal on Selected Areas in Communications
- IEEE Transactions on Communications
- IEEE Transactions on Information Theory
- IEEE Transactions on Vehicular Technology
- IEEE Transactions on Wireless Communications
- Information Sciences, Elsevier

Conference Review

- IEEE International Symposium on Information Theory (ISIT)
- International Symposium on Turbo Codes & Iterative Information Processing (ISTC)
- IEEE Information Theory Workshop (ITW)
- Conference on Neural Information Processing Systems (NIPS)
- IEEE Wireless Communications and Networking Conference (WCNC)

Publications

Journal Papers

(J1) M. Mondelli, S. H. Hassani, and R. Urbanke, "Unified scaling of polar codes: Error exponent, scaling exponent, moderate deviations, and error floors", accepted to *IEEE Trans. Inform. Theory*, Oct. 2016.

(J2) S. Kudekar, S. Kumar, M. Mondelli, H. D. Pfister, E. Şaşoğlu, and R. Urbanke, "Something old, something new, something borrowed, and something proved", *IEEE Inform. Theory Soc. Newslett.*, vol. 65, no. 3, pp. 21–24, Sept. 2015.

(J3) M. Mondelli, S. H. Hassani, and R. Urbanke, "Scaling exponent of list decoders with applications to polar codes", *IEEE Trans. Inform. Theory*, vol. 61, no. 9, pp. 4838–4851, Sept. 2015.

(J4) M. Mondelli, S. H. Hassani, I. Sason, and R. Urbanke, "Achieving Marton's region for broadcast channels using polar codes", *IEEE Trans. Inform. Theory*, vol. 61, no. 2, pp. 783–800, Feb. 2015.

(J5) M. Mondelli, S. H. Hassani, and R. Urbanke, "From polar to Reed-Muller codes: A technique to improve the finite-length performance", *IEEE Trans. Commun.*, vol. 62, no. 9, pp. 3084–3091, Sept. 2014.

(J6) M. Mondelli, Q. Zhou, V. Lottici, and X. Ma, "Joint power allocation and path selection for multi-hop noncoherent decode and forward UWB communications", *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1397–1409, Mar. 2014.

(J7) M. Mondelli, "A finite difference scheme for the stack filter simulating the MCM", *Image Processing On Line*, vol. 3, 2013.

(J8) M. Mondelli and A. Ciomaga, "Finite difference schemes for MCM and AMSS", *Image Processing On Line*, vol. 1, 2011.

Conference Papers

(C1) S. Kudekar, S. Kumar, M. Mondelli, H. D. Pfister, and R. Urbanke, "Comparing the bit-MAP and block-MAP decoding thresholds of Reed-Muller codes on BMS channels", accepted at *IEEE Int. Symposium on Inform. Theory (ISIT)*, Barcelona, Spain, July 2016.

(C2) S. Kudekar, S. Kumar, M. Mondelli, H. D. Pfister, E. Şaşoğlu, and R. Urbanke, "Reed-Muller codes achieve capacity on erasure channels", in *Proc. of the Annual ACM Symposium on Theory of Computing (STOC)*, Boston, MA, USA, June 2016, pp. 658–669. **STOC Best Paper Award**.

(C3) M. Mondelli, S. H. Hassani, and R. Urbanke, "Unified scaling of polar codes: Error exponent, scaling exponent, moderate deviations, and error floors", in *Proc. of the IEEE Int. Symposium on Inform. Theory (ISIT)*, Hong Kong, June 2015, pp. 1422–1426. **IEEE Jack Keil Wolf ISIT Student Paper Award**.

(C4) M. Mondelli, S. H. Hassani, and R. Urbanke, "How to achieve the capacity of asymmetric channels", in *Proc. of the Allerton Conf. on Commun., Control, and Computing*, Monticello, IL, USA, Oct. 2014, pp. 789–796.

(C5) M. Mondelli, S. H. Hassani, I. Sason, and R. Urbanke, "Achieving Marton's region for broadcast channels using polar codes", in *Proc. of the IEEE Int. Symposium on Inform. Theory (ISIT)*, Honolulu, HI, USA, July 2014, pp. 306–310.

(C6) M. Mondelli, S. H. Hassani, and R. Urbanke, "From polar to Reed-Muller codes: A technique to improve the finite-length performance", in *Proc. of the IEEE Int. Symposium on Inform. Theory (ISIT)*, Honolulu, HI, USA, July 2014, pp. 131–135.

(C7) M. Mondelli, S. H. Hassani, and R. Urbanke, "Scaling exponent of list decoders with applications to polar codes", in *Proc. of the IEEE Inform. Theory Workshop (ITW)*, Sevilla, Spain, Sept. 2013, pp. 1–5.

(C8) M. Mondelli, Q. Zhou, X. Ma, and V. Lottici, "A cooperative approach for amplify-and-forward differential transmitted reference IR-UWB relay systems", in *Proc. of the IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Kyoto, Japan, Mar. 2012, pp. 2905–2908.

Preprints / Under Submission

(S1) M. Mondelli, S. H. Hassani, I. Marić, D. Hui, and S.-N. Hong, "Capacity-achieving rate-compatible polar codes for general channels", submitted to *IEEE Wireless Commun. and Networking Conf. (WCNC)*, Nov. 2016.

(S2) M. Mondelli, S. H. Hassani, and R. Urbanke, "How to achieve the capacity of asymmetric channels", submitted to *IEEE Trans. Inform. Theory*, Sept. 2016. [Online]. Available: http://arxiv.org/abs/1406.7373.

(S3) S. Kudekar, S. Kumar, M. Mondelli, H. D. Pfister, E. Şaşoğlu, and R. Urbanke, "Reed-Muller codes achieve capacity on erasure channels", submitted to *IEEE Trans. Inform. Theory*, Jan. 2016. [Online]. Available: http://arxiv.org/abs/1601.04689.

Patents

(P1) M. Mondelli, S. H. Hassani, I. Marić, S.-N. Hong, and D. Hui, "Generalized rate-compatible polar codes", *Ericsson Research*, San Jose, filed in Nov. 2016.

Invited Talks

(T1) "Capacity via symmetry I – A new proof for an old code", *Algorithmic Coding Theory Workshop*, *Institute for Computational and Experimental Research in Mathematics (ICERM)*, Brown University, Providence, June 2016.

(T2) "Capacity via symmetry", *Shannon Centennial Student Competition*, Bell Labs, Nokia, Murray Hill, Apr. 2016.

(T3) "Reed-Muller codes: Thresholds and weight distribution", *International Zurich Seminar on Communications (IZS)*, Zürich, Mar. 2016.

(T4) "Chaining, scaling and Reed-Muller: Two polar paradigms and a conjecture solved", *Graduation Day*, *Information Theory and Applications Workshop (ITA)*, UCSD, San Diego, Feb. 2016.

(T5) "Polar codes: How well they perform and how to make them better", *Ericsson Research*, San Jose, Nov. 2015.

(T6) "Everything you always wanted to know about scaling of polar codes (but were afraid to ask)", *Simons Institute for the Theory of Computing*, Berkeley, Apr. 2015.

(T7) "Unified scaling of polar codes: Error exponent, scaling exponent, moderate deviations, and error floors", *Technische Universität München (TUM)*, Munich, Feb. 2015.

(T8) —, *Graduation Day Poster Session*, *Information Theory and Applications Workshop (ITA)*, UCSD, San Diego, Feb. 2015.

(T9) "Achieving Marton's region for broadcast channels using polar codes", *Conference on Information Sciences and Systems (CISS)*, Princeton University, Mar. 2014.