

POLAR CODES: PERFORMANCE OVER FADING CHANNELS AND  
CONVERGENCE TO REED-MULLER CODES

A THESIS SUBMITTED TO  
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES  
OF  
MIDDLE EAST TECHNICAL UNIVERSITY

BY

IRMAK ÖZVARIŞ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR  
THE DEGREE OF MASTER OF SCIENCE  
IN  
ELECTRICAL AND ELECTRONICS ENGINEERING

MAY 2019



Approval of the thesis:

**POLAR CODES: PERFORMANCE OVER FADING CHANNELS AND  
CONVERGENCE TO REED-MULLER CODES**

submitted by **IRMAK ÖZVARIŞ** in partial fulfillment of the requirements for the degree of **Master of Science in Electrical and Electronics Engineering Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar  
Dean, Graduate School of **Natural and Applied Sciences**

\_\_\_\_\_

Prof. Dr. Tolga Çiloğlu  
Head of Department, **Electrical and Electronics Eng.**

\_\_\_\_\_

Assoc. Prof. Dr. Melek Diker Yücel  
Supervisor, **Electrical and Electronics Eng., METU**

\_\_\_\_\_

**Examining Committee Members:**

Prof. Dr. Erdal Arıkan  
Electrical and Electronics Eng., Bilkent University

\_\_\_\_\_

Assoc. Prof. Dr. Melek Diker Yücel  
Electrical and Electronics Eng., METU

\_\_\_\_\_

Prof. Dr. Çağatay Candan  
Electrical and Electronics Eng., METU

\_\_\_\_\_

Prof. Dr. Elif Uysal Bıyıkoğlu  
Electrical and Electronics Eng., METU

\_\_\_\_\_

Assist. Prof. Dr. Barış Nakiboğlu  
Electrical and Electronics Eng., METU

\_\_\_\_\_

Date: 22.05.2019

**I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.**

Name, Surname: Irmak Özvarış

Signature:

## ABSTRACT

### **POLAR CODES: PERFORMANCE OVER FADING CHANNELS AND CONVERGENCE TO REED-MULLER CODES**

Özvarış, Irmak  
Master of Science, Electrical and Electronics Engineering  
Supervisor: Assoc. Prof. Dr. Melek Diker Yücel

May 2019, 103 pages

Polar codes introduced in 2008 by Erdal Arıkan have been proven to achieve Shannon capacity for any binary-input discrete memoryless channel. Being adopted as a part of the official coding scheme for the 5G standard, up-to-date research has moved from theory to practical applications, albeit keeping the connection with its ancestors. This thesis aims to address these two topics, narrowing down firstly to the performance of polar codes on fading binary symmetric channels and then to the relationship between polar codes and Reed-Muller codes.

For fading channels, we experiment on a hierarchical scheme proposed in 2014 by Si, Köylüoğlu and Viswanath that uses multiple polar coding phases. We simulate the two-state fading case that utilizes three polar codes; two of them designed for binary symmetric channels and one for a binary erasure channel with an erasure rate representing the fading probability. We compare the bit error ratio performance of the proposed scheme with original polar coding. Results show that the hierarchical scheme outperforms the other whenever the probability of being in the degraded channel is not very high.

As for the comparison between polar and Reed-Muller codes, we primarily focus on the generator matrices of the two codes constructed for binary erasure and additive

white Gaussian noise channels. Motivated by the convergence proof of Mondelli; we present some observations asserting the convergence thresholds of polar codes to Reed-Muller codes, in terms of the channel parameters such as erasure probability or signal to noise ratio.

Keywords: Polar Codes, Fading Binary Symmetric Channel, Reed-Muller Codes, Binary Erasure Channel, Binary Input Additive White Gaussian Channel

## ÖZ

### **KUTUPSAL KODLAR: SÖNÜMLEMELİ KANALLARDA PERFORMANS VE REED-MULLER KODLARINA YAKINSAMA**

Özvarış, Irmak  
Yüksek Lisans, Elektrik ve Elektronik Mühendisliği  
Tez Danışmanı: Doç. Dr. Melek Diker Yücel

Mayıs 2019, 103 sayfa

2008 yılında Erdal Arıkan tarafından önerilen kutupsal kodların herhangi ikili-girişli ayrık belleksiz bir kanal için Shannon kapasitesine ulaşabildiği kanıtlanmıştır. 5G standardı resmi kodlama yönteminin bir parçası olarak kabul edilen kodlarla ilgili güncel araştırmalar pratik uygulamalara yoğunlaşsa da bu kodların atalarıyla kuramsal bağlantılarının incelenmesi de sürmektedir. Bu tez; bu iki konuya değinerek, kutupsal kodların sönümlemeli ikili simetrik kanallardaki başarımını ve Reed-Muller kodlarıyla ilişkilerini incelemeyi amaçlamaktadır.

Sönümlemeli kanallar için Si, Köylüoğlu ve Viswanath tarafından 2014 yılında önerilen ve birden fazla kutupsal kodlama fazı içeren hiyerarşik yöntem üzerinde çalışılmıştır. İki durumlu sönümlemeli kanal için; ikisi ikili simetrik kanala, biri de ikili silinti kanalına göre tasarlanmış üç kutup kodu kullanan hiyerarşik kodun benzetimi yapılmıştır. Önerilen yöntemin ikili hata oranı özgün kutupsal kodlamayla karşılaştırıldığında, kötü kanalda kalma olasılığı çok yükselmediği sürece, önerilen yöntemin diğerinden daha iyi bir başarımı olduğu görülmektedir.

Kutup ve Reed-Muller kodları arasındaki karşılaştırmada ise, ikili silinti kanalı ve toplanır beyaz Gauss gürültülü kanal için oluşturulan kutup ve Reed-Muller kodlarının üreten matrislerine odaklanılmıştır. Mondelli'nin yakınsama ispatından yola çıkılarak;

silinti olasılığı veya gürültü varyansı gibi kanal parametreleri cinsinden, kutupsal kodların Reed-Muller kodlarına yakınsama sınırlarını belirten bazı gözlemler sunulmuştur.

Anahtar Kelimeler: Kutupsal Kodlar, Sönümlenmeli İkili Simetrik Kanal, Reed-Muller Kodları, İkili Silinti Kanalı, İkili Girişli Eklenmiş Beyaz Gauss Gürültülü Kanal

To my beloved mom and dad

## ACKNOWLEDGEMENTS

Firstly, I would like to express my heartfelt gratitude to my supervisor, Assoc. Prof. Melek Diker Yücel for providing invaluable guidance throughout this research. Her everlasting passion, motivation and eagerness for any kind of new information have showed me what a good researcher should be. Not only she provided me with professional guidance, but also taught me a great deal in life in general. I cannot express enough thanks for her continued warm-hearted support. I am honored and feel privileged to have worked with her.

My completion of this thesis could not have been accomplished without the support of my colleagues at TÜBİTAK İLTAREN, especially Engin Utku Şenses. His encouragement to complete this thesis has been very much recognized. I would also like to extend my thanks to my friends and my other half Kıvanç who all have been a constant source of support. Their reassurance when the times got rough and sharing my distress are very much appreciated.

Finally, I would like to thank my parents, who always believed in my ability to be successful in whatever I pursue. My mom is my ultimate role model in the academic arena and this research is just the beginning of following her footsteps. Dad, although no longer with us, continues to inspire me by his dedication and devotion to what he believes in. Thank you both for everything and making me who I am right now.

## TABLE OF CONTENTS

ABSTRACT .....	v
ÖZ .....	vii
ACKNOWLEDGEMENTS .....	x
TABLE OF CONTENTS .....	xi
LIST OF TABLES .....	xiii
LIST OF FIGURES .....	xiv
LIST OF ABBREVIATIONS .....	xviii
CHAPTERS	
1. INTRODUCTION .....	1
2. PRELIMINARIES .....	7
2.1. The Discrete Communication Channel .....	7
2.2. The Binary Discrete Memoryless Channel .....	9
2.3. Channel Polarization and Polar Coding .....	12
2.4. Reed Muller Codes .....	24
2.5. The Wireless Fading Channel .....	29
2.6. Further Notes and Literature Review .....	31
2.6.1. Decoders for RM and Polar Codes .....	31
2.6.2. Fading Channels and Polar Codes .....	35
3. HIERARCHICAL POLAR CODING SCHEME FOR FADING BINARY SYMMETRIC CHANNELS.....	37
3.1. Proposed Scheme by Si, Köylüoğlu and Vishwanath .....	37
3.1.1. Hierarchical Encoding .....	40

3.1.1.1. Phase 1 – BEC Encoding .....	40
3.1.1.2. Phase 2 – BSC Encoding .....	41
3.1.2. The Block Fading Channel .....	42
3.1.3. Hierarchical Decoding.....	42
3.1.3.1. Phase 1 – BSC Decoding 1 .....	42
3.1.3.2. Phase 2 – BEC Decoding.....	43
3.1.3.3. Phase 3 – BSC Decoding 2.....	43
3.2. Simulation Results .....	44
4. CHANNEL PARAMETERS AT WHICH POLAR CODES CONVERGE TO REED-MULLER CODES .....	59
4.1. Convergence Proposition by Mondelli .....	60
4.2. BEC Erasure Rates for the Convergence of Polar Codes to RM Codes .....	64
4.3. AWGN Channel SNRs for the Convergence of Polar Codes to RM Codes...	74
4.4. Further Notes on the Convergence with respect to Channel Conditions .....	82
5. DISCUSSION AND CONCLUSION .....	85
REFERENCES .....	91
APPENDICES	
A. Proofs of Some Equations Used .....	99

## LIST OF TABLES

### TABLES

<b>Table 2.1.</b> Channel capacities and initial Bhattacharyya parameters of BSC, BEC and AWGN channels.....	19
<b>Table 3.1.</b> Channel capacities $C_{W_1}, C_{W_2}, C_{\tilde{W}}$ , the overall channel capacity $C$ for $p_1 = 0.1, p_2 = 0.05$ , corresponding number of information bits $K_1, K_2, K_3$ and $K_p$ for $\alpha = 0.6$ .....	47
<b>Table 3.2.</b> Channel capacities $C_{W_1}, C_{W_2}, C_{\tilde{W}}$ , the overall channel capacity $C$ for $p_1 = 0.1, p_2 = 0.02$ , corresponding number of information bits $K_1, K_2, K_3, K_p$ and rate $R_p = 0.6C$ .....	49
<b>Table 3.3.</b> Channel capacities $C_{W_1}, C_{W_2}, C_{\tilde{W}}$ , the overall channel capacity $C$ for $p_1 = 0.1, p_2 = 0.01$ , corresponding number of information bits $K_1, K_2, K_3, K_p$ and rate $R_p = 0.6C$ .....	50
<b>Table 3.4.</b> Channel capacities $C_{W_1}, C_{W_2}, C_{\tilde{W}}$ , the overall channel capacity $C$ for $p_1 = 0.1, p_2 = 0.001$ , corresponding number of information bits $K_1, K_2, K_3, K_p$ and rate $R_p = 0.6C$ .....	52
<b>Table 3.5.</b> Ratio of the polar code rate $R_p$ to capacities of the fading channels $C_{W_1}$ and $C_{W_2}$ for $p_1 = 0.1, p_2 = 0.001$ .....	54
<b>Table 3.6.</b> Channel capacities $C_{W_1}, C_{W_2}, C_{\tilde{W}}$ , the overall channel capacity $C$ for $p_1 = 0.1, p_2 = 0.001$ , corresponding number of information bits $K_1, K_2, K_3, K_p$ and rate $R_p = 0.4C$ .....	56
<b>Table 3.7.</b> BER performance of the BEC decoder of the SKV-code for $p_1 = 0.1, p_2 = 0.001$ , and $\alpha = R/C = 0.4$ .....	57
<b>Table 4.1.</b> Erasure probabilities $\epsilon_0$ and $\epsilon_1$ , for which $(N, N/2) = (2^n, 2^{n-1})$ adaptive polar codes designed for BEC( $\epsilon$ ) converge to RM codes.....	66
<b>Table 4.2.</b> Channel capacities with respect to given SNR values for the AWGN channel.....	79

## LIST OF FIGURES

### FIGURES

<b>Figure 2.1.</b> Block diagram of a digital communication channel (reproduced from [Blahut, 1984]).....	8
<b>Figure 2.2.</b> Block diagram of the simplified digital communication channel .....	9
<b>Figure 2.3.</b> (a) Binary symmetric channel BSC( $p$ ), (b) Equivalent BSC( $p$ ) .....	11
<b>Figure 2.4.</b> Binary erasure channel BEC( $\epsilon$ ) .....	11
<b>Figure 2.5.</b> Additive white Gaussian noise (AWGN) channel. The noise has 0 mean and variance of $\sigma^2$ .....	12
<b>Figure 2.6.</b> Basic channel combining.....	13
<b>Figure 2.7.</b> $W_2^{(1)}$ after basic channel splitting .....	14
<b>Figure 2.8.</b> $W_2^{(2)}$ after basic channel splitting .....	14
<b>Figure 2.9.</b> $W_4$ after recursive channel combining and splitting.....	17
<b>Figure 2.10.</b> Bhattacharyya parameters of (8, 4) polar code designed for BEC(0.5) and selection of the indices (shown by red).....	22
<b>Figure 2.11.</b> Successive cancellation (SC) decoder for polar code of code length 2.... .....	23
<b>Figure 2.12.</b> Belief propagation (BP) decoder construction for RM codes of code length 8. The dashed line shows a processing element (PE) .....	27
<b>Figure 2.13.</b> Details of a processing element (PE) used in a BP decoder.....	28
<b>Figure 3.1.</b> Choice of information bits in a codeword of length N according to the encoding scheme proposed in [Si et al., 2014] for two fading states (reproduced from [Si et al., 2014]). .....	39
<b>Figure 3.2.</b> Polar encoder proposed in [Si et al., 2014] for a fading binary symmetric channel with two states (reproduced from [Si et al., 2014]).....	41
<b>Figure 3.3.</b> Polar decoder proposed in [Si et al., 2014] for a fading binary symmetric channel with two states (reproduced from [Si et al., 2014]).....	45

<b>Figure 3.4.</b> BER performance of the SKV-code and polar codes for different probabilities of the degraded state, $N = B = 256$ , $p_1 = 0.1$ , $p_2 = 0.05$ , with code rates $R_i = 0.6C_i$ .....	48
<b>Figure 3.5.</b> BER performance of the SKV-code and polar codes for different probabilities of the degraded state, $N = B = 256$ , $p_1 = 0.1$ , $p_2 = 0.02$ , with code rates $R_i = 0.6C_i$ .....	50
<b>Figure 3.6.</b> BER performance of the SKV-code and polar codes for different probabilities of the degraded state, $N = B = 256$ , $p_1 = 0.1$ , $p_2 = 0.01$ , with code rates $R_i = 0.6C_i$ .....	51
<b>Figure 3.7.</b> BER performance of the SKV-code and polar codes for different probabilities of the degraded state, $N = B = 256$ , $p_1 = 0.1$ , $p_2 = 0.001$ , with code rates $R_i = 0.6C_i$ .....	53
<b>Figure 3.8.</b> BER performances of the SKV-codes and polar codes for different probabilities of the degraded state, $N = B = 256$ , $p_1 = 0.1$ , $p_2 =$ (a) 0.001, (b) 0.02, (c) 0.01 and (d) 0.001 with code rates $R_i = 0.6C_i$ .....	55
<b>Figure 3.9.</b> BER performance of the SKV-code and polar codes for different probabilities of the degraded state, $N = B = 256$ , $p_1 = 0.1$ , $p_2 = 0.001$ , with code rates $R_i = 0.4C_i$ .....	57
<b>Figure 3.10.</b> BER performances of the SKV-code and polar codes for different probabilities of the degraded state, $N = B = 256$ , $p_1 = 0.1$ , $p_2 = 0.001$ , with code rates $R_i = 0.6C_i$ . Additionally, (a) polar codes are designed with respect to the degraded channel $W_1$ , and (b) to the superior channel $W_2$ , at rates $R_p, R_1$ and $R_2$ ....	58
<b>Figure 4.1.</b> Polar code construction for $N = 2$ . .....	62
<b>Figure 4.2.</b> Polar code construction for $N = 4$ . .....	63
<b>Figure 4.3.</b> Variation of design erasure probabilities $\epsilon_0$ and $\epsilon_1$ versus $n$ , for which $(2^n, 2^{n-1})$ adaptive polar codes designed for BEC( $\epsilon$ ) converge to RM codes. ....	65
<b>Figure 4.4.</b> Number of different generator matrix rows $D(n, 0.5, \epsilon)$ of the $(2^n, 2^{n-1})$ RM and adaptive polar codes designed for BEC( $\epsilon$ ) versus $n$ , for various design erasure rates. ....	67

**Figure 4.5.** Normalized ratio of different rows,  $D(n, 0.5, \epsilon)/K$ , between the generator matrices of the  $(2^n, 2^{n-1})$  RM and adaptive polar codes designed for BEC( $\epsilon$ ) versus  $n$ , for various erasure rates. .... 68

**Figure 4.6.** Number of differing rows  $D(n, R, 0.1)$  between the generator matrices of the adaptive polar codes designed for BEC(0.1) and RM codes at various rates. .... 69

**Figure 4.7.** Normalized number  $D(n, R, 0.1)/K$  of differing rows between the generator matrices of the adaptive polar codes designed for BEC(0.1) and RM codes, versus the code rate  $R$ . .... 71

**Figure 4.8.** Normalized number  $D(n, R, \epsilon)/K$  of differing rows between the generator matrices of the adaptive polar codes designed for BEC( $\epsilon$ ) and RM codes, versus the code rate  $R$ . .... 73

**Figure 4.9.** Variation of the design SNR values versus  $n$ , for which  $(2^n, 2^{n-1})$  adaptive polar codes converge to RM codes. .... 75

**Figure 4.10.** Number of different generator matrix rows  $D'(n, 0.5, \text{SNR})$  of the  $(2^n, 2^{n-1})$  RM and adaptive polar codes designed for an AWGN at that SNR, versus  $n$ , for various design SNRs. .... 76

**Figure 4.11.** Normalized number of different rows  $D'(n, 0.5, \text{SNR})/K$ , between the generator matrices of the  $(2^n, 2^{n-1})$  RM and adaptive polar codes versus  $n$ , for various polar code design SNRs. .... 77

**Figure 4.12.** Number of differing rows  $D'(n, R, 0)$  between the generator matrices of the adaptive polar codes (designed for AWGN at SNR = 0 dB) and RM codes, versus rate  $R$ . .... 78

**Figure 4.13.** Normalized number  $D'(n, R, 0)/K$  of differing rows between the generator matrices of the adaptive polar codes designed for AWGN channel with SNR = 0 dB and RM codes, versus the code rate  $R$ . .... 79

**Figure 4.14.** Normalized number  $D'(n, R, \text{SNR})/K$  of differing rows between the generator matrices of the adaptive polar codes designed for AWGN channel and RM codes, versus the code rate  $R$ , at various SNRs. .... 81

<b>Figure 4.15.</b> Normalized number of different rows $D(n, 0.5, \epsilon)/K$ , between the generator matrices of the $(2^n, 2^{n-1})$ RM and adaptive polar codes designed for that $\epsilon$ , versus the design $\epsilon$ .....	83
<b>Figure 4.16.</b> Normalized number of different rows $D'(n, 0.5, SNR)/K$ , between the generator matrices of the $(2^n, 2^{n-1})$ RM and adaptive polar codes designed for that $SNR$ , versus the design $SNR$ of the adaptive polar code. ....	83
<b>Figure 4.17.</b> Normalized number of different rows $D(n, R, SNR)/K$ , between the generator matrices of the RM and adaptive polar codes designed for that $SNR$ , versus the design $SNR$ of the adaptive polar code at rates $R = 0.3, 0.5$ and $0.7$ .....	84
<b>Figure A.1.</b> Probability distribution function of $p_Y(y x)$ .....	99

## LIST OF ABBREVIATIONS

### ABBREVIATIONS

AWGN	Additive White Gaussian Noise
B-DMC	Binary Input Discrete Memoryless Channel
BEC	Binary Erasure Channel
BER	Bit Error Ratio
BI-AWGN	Binary Input Additive White Gaussian Noise
BMS	Binary Memoryless Symmetric
BP	Belief Propagation
BPL	Belief Propagation List
BSC	Binary Symmetric Channel
CSI	Channel State Information
LLR	Log Likelihood Ratio
RM	Reed-Muller
SC	Successive Cancellation
SCL	Successive Cancellation List
SKV	Si, Köylüoğlu, Vishwanath (of [Si, Köylüoğlu, & Vishwanath, 2014])
SNR	Signal-to-Noise Ratio

## CHAPTER 1

### INTRODUCTION

Coding theory dates back to 1948, when Shannon published his seminal paper [Shannon, 1948], posing the fundamental problem for a communication system of how a message of information can be transmitted efficiently and reliably across a noisy channel. Although he gave the answer to this problem as “coding” and proved that such suitable codes exist, he did not specifically address how to find these good codes. Since then, finding the structures for practical coding schemes that approach Shannon’s theoretical limits has been one of the main focuses of research in information theory and communications.

As formalized by Shannon, the aforementioned problem can be divided into two separate problems: the *source coding* and the *channel coding*. The former’s task is to efficiently represent the source of information using least possible number of bits; while the latter adds redundancy to protect the information against the noisy channel and reliably transmit them to the receiver. In this thesis, our focus is on the channel coding, and it is assumed that the source coding problem has already been solved.

The most important parameters for the transmission problem are rate, probability of (block or bit) error, delay and (encoding or decoding) complexity [Richardson & Urbanke, 2008]. According to Shannon’s channel coding theorem, all rates below capacity  $C$  are achievable. That is, for every rate  $R < C$ , there exists a sequence of codes with maximal probability of error approaching to zero as codelength goes to infinity. Conversely, any sequence of codes with the maximal probability of error approaching zero must have a rate  $R \leq C$ . From this theorem, it is evident that we can and would want to send our information at high rates with low probability of error. But as the codelength increases, the constraints on delay and complexity begin to rise.

Numerous research has been done to balance this trade-off, along with the explicit constructions of the codes with arbitrarily small probability of error. Hamming introduced the first single-error-correcting block codes in 1950, which was quite weak with respect to what Shannon proposed, yet was the best code discovered in that time [Hamming, 1950]. *Hamming codes* were then followed by the binary and ternary *Golay codes* introduced by Golay [Golay, 1949], multiple error correcting *Reed-Muller codes* by Reed and Muller [Muller, 1954; Reed, 1954], which happen to be a close relative to polar codes that will be discussed later, and the *BCH codes* by Bose and Ray [Bose & Ray-Chaudhuri, 1960], and independently by Hocquenghem (Hocquenghem, 1959). Around the same time, Reed and Solomon proposed *Reed-Solomon codes* as a special case of BCH codes [Reed & Solomon, 1960]. By 1970s, the algebraic approach whose objective was to maximize the minimum distance to maximize the error correction radius, left its seat to the probabilistic approach which is concerned with optimizing the performance with respect to the encoding and decoding complexity with the invention of *convolutional codes* by Elias [Elias, 1955], *sequential decoding* by Wozencraft [Wozencraft, 1957] and further development of Elias' work by Forney [G. D. Forney, 1970]. The first generation of channel coding technology then started with introduction of the iterative decoding algorithms such as *Viterbi algorithm* by Viterbi [Viterbi, 1967] and the *BCJR algorithm* by Bahl, Cocke, Jelinek, and Raviv [Bahl, Cocke, Jelinek, & Raviv, 1974]. The Viterbi algorithm was adopted for the 2G GSM networks. In 1993, the second generation of channel coding technology began with the epoch-making *turbo codes* that were designed by Berrou, Glavieux and Thitimajshima [Berrou, Glavieux, & Thitimajshima, 1993] having a good error performance, rate very close to the Shannon capacity and linear decoding complexity, and adopted by 3GPP for the 3G UMTS systems. A couple of years later, MacKay rediscovered the *LDPC codes* [MacKay & Neal, 2002] with the comparable properties to the turbo codes, which was originally developed by Gallager way back in 1962 [Gallager, 1962].

Both Turbo and LDPC codes belong to the family of *codes on graphs* that again goes back to 1981 when Tanner founded this capacity-achieving field of codes with the iterative decoding algorithm [Tanner, 1981]. These two codes are extensively used in communication standards: Turbo being used in deep space communications (CCSDS), and in the 3G/4G standards; LDPC being used for digital video broadcasting, satellite communications, IEEE 802.3an (Ethernet), 802.16e (WiMax) 802.11n/ac (WiFi), recent mobile generation (4G, LTE) and the next generation (5G).

Although it may seem that the problem imposed by Shannon has been already solved, as Costello and Forney summarized in [Costello & Forney, 2007], “Coding is not dead” yet. With the introduction of a new coding scheme called as *polar codes* by Arıkan [Arıkan, 2009], the third generation of channel coding technology has started and both LDPC and polar codes have since been adopted for the 5G. Polar codes are based on theoretically proven *channel polarization* phenomenon where  $N$  independent copies of the channel are combined and split so that the overall channels *polarize* in the sense that some portion of the channel indices’ capacities tend to 1; i.e. the channels become purely noiseless, while some tend to 0; i.e. the channels become purely noisy as  $N \rightarrow \infty$ . With this effect, one can reliably send information over the noiseless channels, while sending known bits, or in other words *frozen bits*, over the noisy channels. In the decoder, Arıkan originally uses *successive cancellation (SC) decoding* that has a time complexity of  $O(N \log N)$ . When compared to Turbo and LDPC code performances, although this decoding scheme has poor performance, with other decoding schemes like belief propagation (BP) or list decoding of SC (SCL), it is shown by Tal and Vardy that polar codes indeed outperform Turbo and LDPC code performances [Tal & Vardy, 2015]. In short, this coding scheme provably achieves the theoretical channel capacity with low complexity for arbitrary symmetric discrete memoryless channels, both with a binary (B-DMC) [Hussami, Korada, & Urbanke, 2009] and non-binary input alphabet [Tal & Vardy, 2013].

Countless research has been done since the invention of polar codes to make the performance better, including new decoding techniques, concatenation with other schemes, and also some hardware implementations. A brief information about the work is given in Chapter 2, albeit it is impossible to include all aspects of rapidly growing current research. An interesting area is the design of polar codes for the wireless communication channels, where fading aspects and time-varying nature of the channel presents major difficulty. Wireless communication has become one of the most dynamic area of research in recent years due to increasing demand in not only daily-used appliances such as cellular phones, tablets, laptops, or generally speaking, remotely controlled devices; but also larger-scale needs such as smart homes that use intelligent home electronics, satellite communications, commercial and military applications. Evidently, this demand brings proliferating research seeking for better, newer and more reliable technologies. Since the polar code performance has already been theoretically proved, this low-complexity-coding scheme is quite compelling for the developing systems, such as the fifth-generation (5G) cellular services. In fact, in 2006, the third generation partnership project (3GPP), which is the international standards organization that develops telecommunication protocols, has adopted polar codes for high data rate demanding control channels of the air interface for 5G. However, as the main propagation in wireless communication is electromagnetic wave propagation in air, there are many unpredicted and random limitations, such as moving objects or simply weather, as well as the interference due to vastly increasing number of users. The prior creates variation in signal power over time and frequency and is called as *fading*. In a fading channel, there are replicas of the signal with different amplitudes, phases and angles of arrival. These replicas are called as *multipaths* which may add up constructively or destructively at the receiver. This multipath propagation environment changes in a random manner and as a result, understanding the random behavior and proposing solutions to overcome the adverse impacts of fading have been among the key aspects of wireless communication channels.

Furthermore, Arikan points out the relationship between polar and other codes such as RM and BCH codes as early as 2009, in [Arikan, 2009]. It is stated that the similarities between the code constructions of polar and RM codes became clear such that they belong to the same class of codes and the only difference is the rule for selecting the rows from a Hadamard matrix as the basis vectors of their generator matrices. In addition, another similarity between RM and polar codes is that they both can be constructed using Plotkin's construction: The construction of RM codes starts with smaller codes and recursively grows, while that of polar starts from the full-order generator matrix and removes the unnecessary rows. Without going further, Arikan also interprets polar codes in a spectral point of view [Arikan, 2009] which is also mentioned for BCH codes by Blahut [Blahut, 1984]. In short, it is intriguing to see how a (relatively) new born code has relationship and similarities with two of the earliest codes. Due to simplicity and regarding the previous work of Akdoğan [Akdoğan, 2018], we choose to examine the relationship between RM and polar codes in this study.

The thesis consists of two independent parts: In the first part, it is aimed to observe the performance of polar codes in a block fading channel, comparing the performance of the hierarchical scheme [Si, Köylüoğlu, & Vishwanath, 2014] that uses multiple polar coding phases to that of Arikan's plain polar coding scheme [Arikan, 2009]. In the second part, it is aimed to calculate numerically the convergence of polar codes to RM codes, specifically for the BEC and AWGN channel. The organization of this thesis is as follows:

In Chapter 2, preliminary information about various concepts such as communication channels, coding and information theoretic parameters are given. Channel polarization is summarized and encoding and decoding structures of both polar and its close relative Reed-Muller codes are explained. The wireless fading channels are reviewed in general terms, and the chapter is finalized with a review on up-to-date work relating to decoders and also polar codes on fading channels.

In Chapter 3, BER performance of the hierarchical polar coding scheme for binary fading channels proposed by Si, Köylüoğlu and Vishwanath [Si et al., 2014] and that of Arikan's original polar code scheme is compared. The proposed scheme is reviewed in detail and simulation results under various conditions are given.

In Chapter 4, the convergence of polar codes to RM codes is discussed with numerical results. The proposition relating to the convergence given by Mondelli [Mondelli, 2016] is reviewed in detail. Some numerical values of design erasure rates or design SNRs respectively, required by the convergence of polar codes to RM codes, for the BEC and the AWGN channel are calculated, and various cases showing the convergence are presented.

In Chapter 5, concluding remarks and contributions of this thesis are discussed.

## CHAPTER 2

### PRELIMINARIES

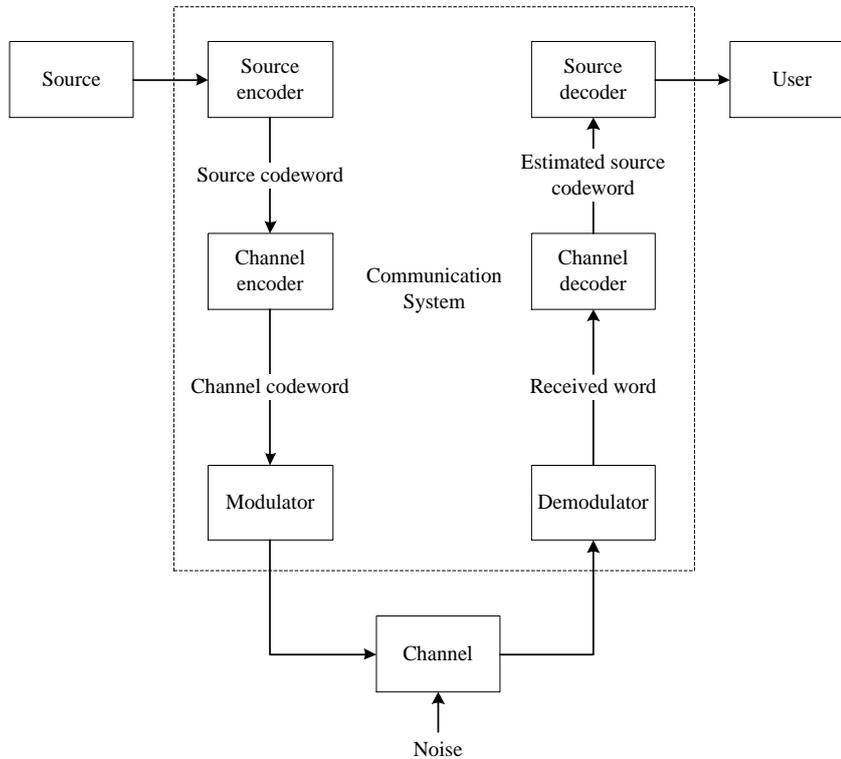
In this chapter, we give some preliminaries that are used in the thesis. We start with the discrete communication channel models, followed by the phenomenon of channel polarization based on Arıkan's work [Arıkan, 2009] with reference to Korada's doctoral thesis [Korada, 2009], and Reed Muller codes. Continuing with a brief description of wireless fading channels, we end the chapter by giving a brief survey on other recent work regarding decoders and fading channels within the context of this thesis.

The notation throughout the thesis is as follows: The capital letters  $X, Y$  denote the random variables whose sample values are the lowercase ones  $x, y$ . We write  $X^N$  to denote the random variables  $(X_1, X_2, \dots, X_N)$  and  $X_i^j: 1 \leq i, j \leq N$  to denote the sub-vector  $(X_i, X_{i+1}, \dots, X_j)$ . The same analogy is used for  $x^N$  and  $x_i^j$ , as well. We denote the channel as  $W: \mathcal{X} \rightarrow \mathcal{Y}$  with  $\mathcal{X}$  and  $\mathcal{Y}$  being the input and output alphabets, respectively. A bold  $\mathbf{x}$  is used to denote vectors and matrices such that  $\mathbf{x} =$

$$\begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \text{ in general.}$$

#### 2.1. The Discrete Communication Channel

As discussed before, the main goal of a channel code is to transmit data reliably and efficiently through the communication channel. Shannon declared that this was possible; but as he did not say how, we are interested in finding such codes. To begin with, we start with the basic digital communication system:

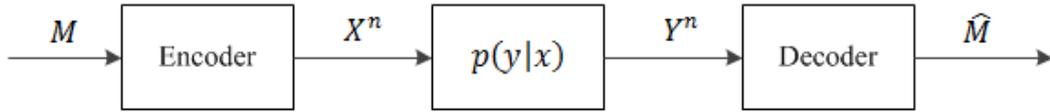


**Figure 2.1.** Block diagram of a digital communication channel (reproduced from [Blahut, 1984])

Throughout the thesis, as mentioned in previous chapter, we are only concerned with the channel encoder and decoder, which is known as channel coding. Our communication system is then depicted in Figure 2.2. The channel is a discrete memoryless channel (DMC) consisting of an encoding function that maps each message  $m \in \mathcal{M}$  to codewords  $x^N \in \mathcal{X}^N$ , forming the codebook, and a decoding function that tries to find the estimates  $\hat{m}$  of  $m$  from the received sequence  $y^N \in \mathcal{Y}^N$ . We assume that the message is uniformly distributed. The channel is memoryless in the sense that  $p(y^N|x^N, m) = \prod_{i=1}^N p(y_i|x_i)$  where  $p(y|x)$  denotes the conditional probability of receiving the output  $y \in \mathcal{Y}$  and  $N$  denotes the number of channel uses. The performance of the code is measured by the probability of making an error, i.e. probability of the estimate  $\hat{m}$  not being equal to  $m$ . A rate is said to be achievable if

there exists a sequence of codes such that average probability of error approaches 0 which is the consequence of Shannon's Coding Theorem and its converse; and then, the capacity  $C$  of a DMC is the supremum over all achievable rates [Tse & Viswanath, 2005]:

$$C = \max_{p(x)} I(X; Y) \quad (2.1)$$



**Figure 2.2.** Block diagram of the simplified digital communication channel

## 2.2. The Binary Discrete Memoryless Channel

In this thesis, the input alphabet  $\mathcal{X}$  is binary, so we have the binary-input discrete memoryless channel (B-DMC). All of the computations are in modulo-2, the logarithms are to the base 2 (unless stated otherwise) and the channel capacities and code rates are in bits. If we denote the B-DMC as  $W$ , we can write  $W: \mathcal{X} \rightarrow \mathcal{Y}$  with the conditional probability (denoted as  $p$  above in Fig. 2.2)  $W(y|x)$ , and  $W^N$  corresponds to  $N$  uses of the channel; i.e. we have  $W^N: \mathcal{X}^N \rightarrow \mathcal{Y}^N$  with  $W(y^N|x^N, m) = \prod_{i=1}^N W(y_i|x_i)$  in correspondence to aforementioned general DMC.

For a B-DMC, two parameters of interest are the symmetric capacity:

$$I(W) \triangleq \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} \frac{1}{2} W(y|x) \log \frac{W(y|x)}{\frac{1}{2} W(y|0) + \frac{1}{2} W(y|1)} \quad (2.2)$$

and the Bhattacharyya parameter:

$$Z(W) \triangleq \sum_{y \in \mathcal{Y}} \sqrt{W(y|0)W(y|1)} \quad (2.3)$$

Although  $I(\cdot)$  denotes the mutual information between the input  $X$  and output  $Y$ , from (2.1), we can also call this as the symmetric capacity.  $I(W)$  provides a measure of rate: It is the highest rate at which one can make a reliable communication for a symmetric channel.  $Z(W)$  is the upper bound on the probability of maximum-likelihood (ML) decision error and provides a measure of reliability [Arıkan, 2009]. Both of these parameters take values in the closed interval  $[0, 1]$ . The importance of both parameters will show up in designing the polar codes which will be discussed next, thus we will not be giving further details about these two here, except the following relationship between them:

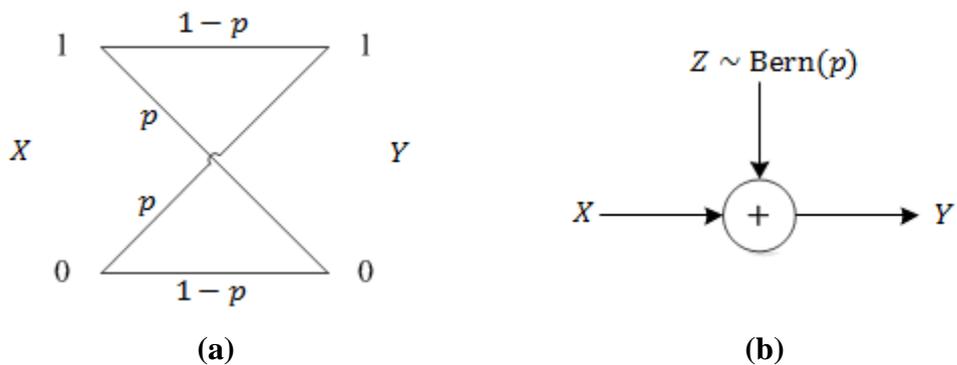
$$I(W) + Z(W) \geq 1 \quad (2.4)$$

$$I(W) \leq \sqrt{1 - Z(W)^2} \quad (2.5)$$

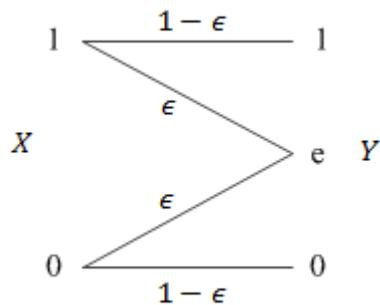
We can interpret (2.4) and (2.5) as higher the capacity, lower the Bhattacharyya parameter. The proof is given in [Arıkan, 2012].

$I(W)$  is equal to the Shannon capacity when  $W$  is a symmetric channel [Arıkan, 2009]. In this thesis, we use the following binary memoryless channels: The binary symmetric channel (BSC), binary erasure channel (BEC) and additive white Gaussian noise (AWGN) channel. BSC is symmetric in the sense that  $W(0|0) = W(1|1)$  and  $W(1|0) = W(0|1) = p$ , i.e. the probability of making an error is  $p$ , which is called as the transmission or the crossover probability. For BEC, we have  $W(y|0)W(y|1) = 0$ , and an error is made when the information bit is erased. In this case, we have  $W(y|0) = W(y|1) = \epsilon$  and the information bit is erased with probability  $\epsilon$ . We denote these two channels as BSC( $p$ ) and BEC( $\epsilon$ ), respectively, which are shown in Figure 2.3(a) and 2.4. BSC also represents a discrete additive noise channel, where the noise has Bernoulli distribution with the same parameter  $p$  as shown in Figure 2.3(b). We will use this property in the next chapter, thus give the proof hereinafter.

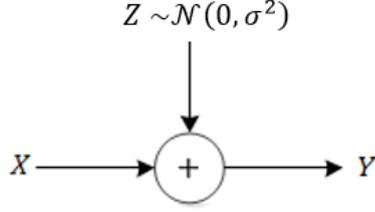
An AWGN channel is similar to as shown in Figure 2.3(b), but instead of a Bernoulli( $p$ ) distributed noise, we have a white Gaussian noise. The noise is called “white” because the power is constant for all frequencies; i.e. its spectral density is constant, and it has a “Gaussian” distribution. The capacity of BSC( $p$ ) is  $1 - H(p)$  with  $H(\cdot)$  being the binary entropy function, that of BEC( $\epsilon$ ) is simply  $1 - \epsilon$  and that of an AWGN channel is  $\frac{1}{2} \log_2(1 + \text{SNR})$  where SNR is the signal-to-noise ratio.



**Figure 2.3.** (a) Binary symmetric channel BSC( $p$ ), (b) Equivalent BSC( $p$ )



**Figure 2.4.** Binary erasure channel BEC( $\epsilon$ )



**Figure 2.5.** Additive white Gaussian noise (AWGN) channel. The noise is 0 mean and of variance  $\sigma^2$

### 2.3. Channel Polarization and Polar Coding

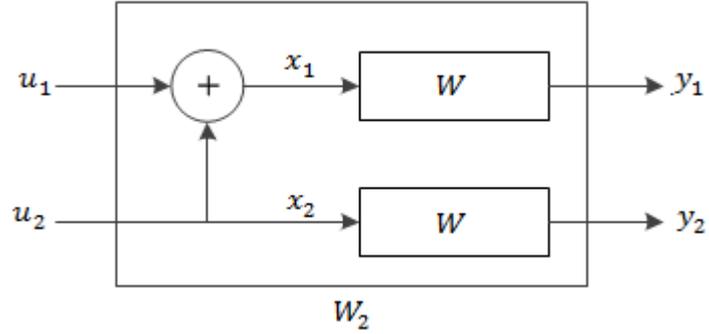
Channel polarization occurs as a result of (i) combining and (ii) splitting  $N$  independent copies of B-DMCs  $W$ , forming  $N$  binary-input channels  $W_N^{(i)}$ :  $1 \leq i \leq N$  where almost each of the newly constructed channels' capacities approach to the two extremes, 0 and 1 as  $N$  goes to infinity. Formal proof of this phenomenon is given in [Arkan, 2009, Theorem 1]. We will start with the first level of polarization, and then continue to the general case.

We start with the basic channel transform using two individual channels  $W$ . In the first step of the recursion,  $N = 2$  and  $W_2: \mathcal{X}^2 \rightarrow \mathcal{Y}^2$  is obtained as shown in Figure 2.6. The newly constructed channel's transition probability is given as

$$W_2(y_1^2 | u_1^2) \triangleq \prod_{i=1}^2 W(y_i | x_i) = W(y_1 | u_1 \oplus u_2) W(y_2 | u_2) \quad (2.6)$$

Since the linear transform between  $U_1^2$  and  $X_1^2$  is one-to-one (mod-2 sum), and  $U_1^2$  being identically independent distributed (i.i.d.) implies  $X_1^2$  also being i.i.d., we have the following:

$$I(U_1^2; Y_1^2) = I(X_1^2; Y_1^2) = I(X_1; Y_1) + I(X_2; Y_2) = 2I(W) \quad (2.7)$$



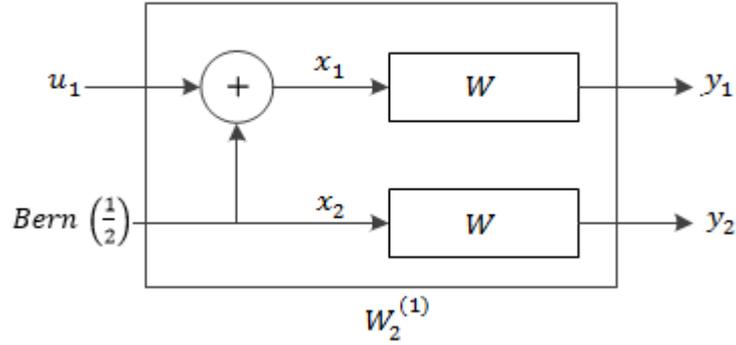
**Figure 2.6.** Basic channel combining

(2.7) implies that the channel capacity is conserved under the single-step transform. Up to now, this phase is called as *channel combining* [Arikan, 2009]. We can split left hand side of (2.7) by using the chain rule:

$$\begin{aligned}
 I(U_1^2; Y_1^2) &= I(U_1; Y_1^2) + I(U_2; Y_1^2 | U_1) \\
 &= I(U_1; Y_1^2) + I(U_2; Y_1^2, U_1)
 \end{aligned} \tag{2.8}$$

(2.8) can be interpreted as follows: First term of the right hand side,  $I(U_1; Y_1^2)$ , is the mutual information between the input  $U_1$  and the outputs  $Y_1$  and  $Y_2$  with  $U_2$  treated as random (noise) as shown in Figure 2.7. Denoting this channel as  $W_2^{(1)}: \mathcal{X} \rightarrow \mathcal{Y}^2$ , by marginalizing (2.6) over  $U_2$ , the transition probability is given as:

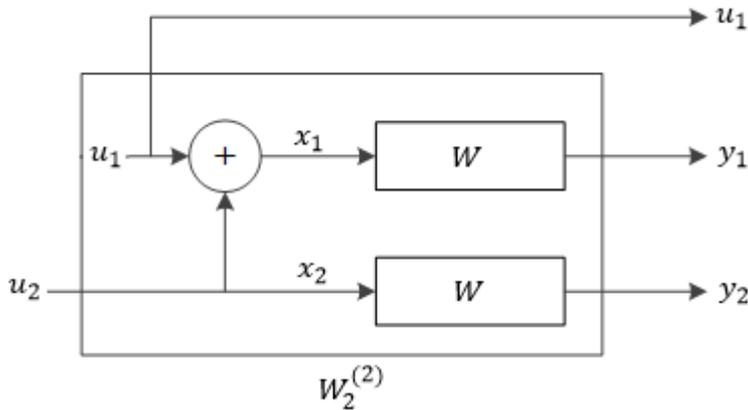
$$\begin{aligned}
 W_2^{(1)}(y_1^2 | u_1) &= \frac{1}{2} \sum_{u_2} W_2(y_1^2 | u_1^2) \\
 &= \frac{1}{2} \sum_{u_2} W(y_1 | u_1 \oplus u_2) W(y_2 | u_2)
 \end{aligned} \tag{2.9}$$



**Figure 2.7.**  $W_2^{(1)}$  after basic channel splitting

Next, second term of the right-hand side of (2.8),  $I(U_2; Y_1^2, U_1)$ , is the mutual information between the input  $U_2$  and the outputs  $Y_1$  and  $Y_2$  with  $U_1$  known as shown in Figure 2.8. Similarly, denoting this channel as  $W_2^{(2)}: \mathcal{X} \rightarrow \mathcal{Y}^2 \times \mathcal{X}$ , the transition probability is given as:

$$W_2^{(2)}(y_1^2, u_1 | u_2) = \frac{1}{2} W_2(y_1^2 | u_1^2) = \frac{1}{2} W(y_1 | u_1 \oplus u_2) W(y_2 | u_2) \quad (2.10)$$



**Figure 2.8.**  $W_2^{(2)}$  after basic channel splitting

With (2.9) and (2.10) one can see that the channel is split into two sub-channels,  $W_2^{(1)}$  and  $W_2^{(2)}$ . Arkan calls this phase as *channel splitting* [Arkan, 2009]. With these two phases, the polarization effect can be seen from the basic channel transform as follows: Combining (2.7), (2.8), (2.9) we have

$$I(W_2^{(1)}) + I(W_2^{(2)}) = 2I(W) \quad (2.11)$$

and further,

$$\begin{aligned} I(W_2^{(2)}) &= I(U_2; Y_1^2, U_1) \\ &= H(U_2) - H(U_2 | Y_1^2, U_1) \\ &\geq H(U_2) - H(U_2 | Y_2) = I(W) \end{aligned} \quad (2.12)$$

This concludes that

$$I(W_2^{(1)}) \leq I(W) \leq I(W_2^{(2)}) \quad (2.13)$$

Likewise, the Bhattacharyya parameter is transformed as

$$Z(W_2^{(1)}) \leq 2Z(W) - Z(W)^2 \quad (2.14)$$

$$Z(W_2^{(2)}) = Z(W)^2 \quad (2.15)$$

The proof of (2.14) and (2.15) is given in [Korada, 2009]. Combining these two, we have

$$Z(W_2^{(1)}) + Z(W_2^{(2)}) \leq 2Z(W) \quad (2.16)$$

$$Z(W_2^{(1)}) \geq Z(W) \geq Z(W_2^{(2)}) \quad (2.17)$$

(2.11-17) imply that instead of using two independent channels, by combining and splitting them, we get one better and one worse channel (i.e. two channels with either higher capacity or lower capacity, respectively (or with lower and higher

Bhattacharyya parameters, respectively). Due to this, we can also denote  $W_2^{(1)}$  and  $W_2^{(2)}$  as  $W^-$  and  $W^+$ , respectively, of the initial channel  $W$ . Once we have such channels; intuitively, we want to send the information from the better channel, while sending a known variable - Arıkan calls these known variables as *frozen bits* - from the bad channels. This is the main notion in construction of polar codes.

With the basic transform having discussed, we can move onto the general case starting with the second step of recursion where  $N = 4$  for illustrative purposes. For  $N = 4$ , similar to combining two independent copies of  $W$  to form  $W_2$ , two independent copies of  $W_2$  are combined to obtain  $W_4$ , with the transition probability

$$\begin{aligned} W_4(y_1^4|u_1^4) &\triangleq \prod_{i=1}^4 W(y_i|x_i) \\ &= W(y_1|u_1 \oplus u_2 \oplus u_3 \oplus u_4)W(y_2|u_2 \oplus u_4)W(y_3|u_3 \oplus u_4)W(y_4|u_4) \quad (2.18) \\ &= W_2(y_1^2|u_1 \oplus u_3, u_2 \oplus u_4)W_2(y_3^2|u_3, u_4) \end{aligned}$$

The combining phase of  $N = 4$  is illustrated in Figure 2.9. In general, we can write the following transition probabilities for the newly constructed channel  $W_N: \mathcal{X}^N \rightarrow \mathcal{Y}^N$  formed by the combining and splitting operations. Combining phase yields

$$W_N(y^N|u^N) \triangleq W_{N/2}(y_1^{N/2}|u_o^N \oplus u_e^N)W_{N/2}(y_{N/2+1}^N|u_e^N) \quad (2.19)$$

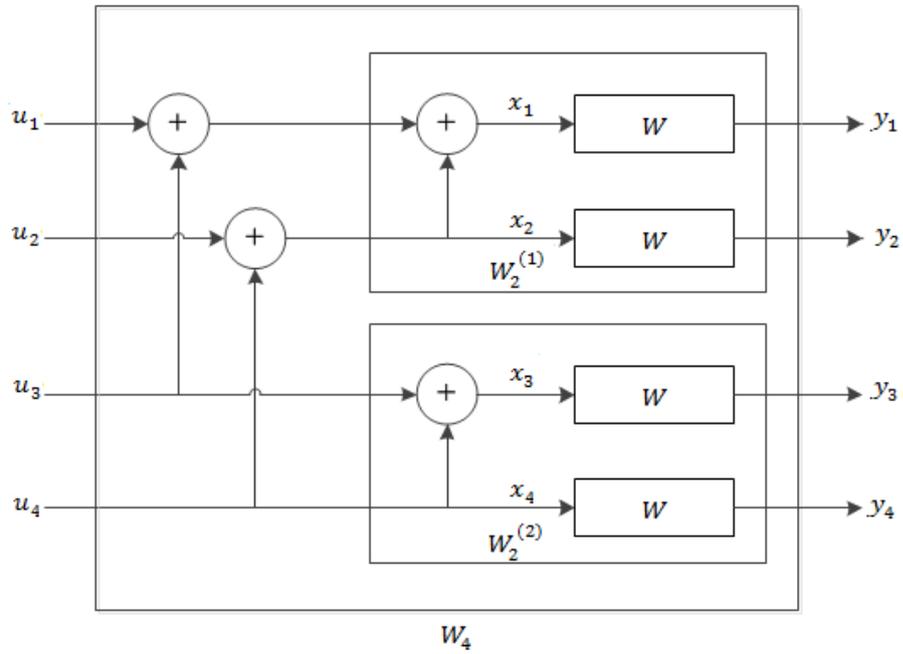
where  $u_o^N$  are the  $u_i$ 's with odd indices, i.e.  $u_o^N = \{u_1, u_3, u_5, \dots, u_{N-1}\}$ , and  $u_e^N$ 's are the even indices, i.e.  $u_e^N = \{u_2, u_4, u_6, \dots, u_N\}$ . For  $1 \leq i \leq N$ , the splitting phase yields

$$W_N^{(i)}(y_1^N, u_1^{i-1}|u_i) \triangleq \sum_{u_{i+1}^N \in \mathcal{X}^{N-i}} \frac{1}{2^{N-i}} W_N(y^N|u^N) \quad (2.20)$$

with the relation between  $W_N^{(i)}$  and the previously constructed  $W_{N/2}^{(j)}$  channels where  $1 \leq j \leq N/2$  can be written as

$$\begin{aligned}
& W_N^{(2j-1)}(y_1^N, u_1^{2j-2} | u_{2j-1}) \\
&= \frac{1}{2} \sum_{u_{2j}} W_{\frac{N}{2}}^{(j)} \left( y_1^{\frac{N}{2}}, u_{1,o}^{2j-2} \oplus u_{1,e}^{2j-2} \middle| u_{2j-1} \oplus u_{2j} \right) W_{\frac{N}{2}}^{(j)}(y_{\frac{N}{2}+1}^N, u_{1,e}^{2j-2} | u_{2j})
\end{aligned} \tag{2.21}$$

$$\begin{aligned}
& W_N^{(2j)}(y_1^N, u_1^{2j-1} | u_{2j}) \\
&= \frac{1}{2} W_{\frac{N}{2}}^{(j)} \left( y_1^{\frac{N}{2}}, u_{1,o}^{2j-2} \oplus u_{1,e}^{2j-2} \middle| u_{2j-1} \oplus u_{2j} \right) W_{\frac{N}{2}}^{(j)}(y_{\frac{N}{2}+1}^N, u_{1,e}^{2j-2} | u_{2j})
\end{aligned} \tag{2.22}$$



**Figure 2.9.**  $W_4$  after recursive channel combining and splitting

Furthermore, rate and reliability parameters are transformed as

$$I(W_N^{(2j-1)}) \leq I(W_{\frac{N}{2}}^{(j)}) \leq I(W_N^{(2j)}) \tag{2.23}$$

$$\left(I_N^{(2j-1)}\right) + I\left(W_N^{(2j)}\right) = 2I\left(W_{N/2}^{(j)}\right)$$

and

$$\begin{aligned} Z\left(W_N^{(2j-1)}\right) &\leq 2Z\left(W_{N/2}^{(j)}\right) - Z\left(W_{N/2}^{(j)}\right)^2 \\ Z\left(W_N^{(2j)}\right) &= Z\left(W_{N/2}^{(j)}\right)^2 \\ Z\left(W_N^{(2j-1)}\right) + Z\left(W_N^{(2j)}\right) &\leq 2Z\left(W_{N/2}^{(j)}\right) \\ Z\left(W_N^{(2j-1)}\right) &\geq Z\left(W_{N/2}^{(j)}\right) \geq Z\left(W_N^{(2j)}\right) \end{aligned} \tag{2.24}$$

Similar to (2.7), the cumulative rate and reliability is given as

$$\begin{aligned} \sum_{i=1}^N I\left(W_N^{(i)}\right) &= NI(W) \\ \sum_{i=1}^N Z\left(W_N^{(i)}\right) &\leq NZ(W) \end{aligned} \tag{2.25}$$

More detail and proof can be found in [Arıkan, 2009; Korada, 2009]. Equations regarding the Bhattacharyya parameter hold with equality if and only if  $W$  is BEC.

Starting from the initial Bhattacharyya parameter denoted as  $Z_{1,1}$ , one can calculate the Bhattacharyya parameters recursively as

$$Z_{2k,j} = \begin{cases} 2Z_{k,j} - Z_{k,j}^2 & , 1 \leq j \leq k \\ Z_{k,j}^2 & , k + 1 \leq j \leq 2k \end{cases}$$

for  $k = 1, 2, 2^2, \dots, 2^{n-1}$ . For the channels mentioned above in Figures 2.3 – 2.5, the initial Bhattacharyya parameters are summarized in Table 2.1, along with the channel capacities. Since the initial Bhattacharyya parameter of an AWGN channel is a function of the noise variance  $\sigma^2$  [H. Li & Yuan, 2013], we will denote it as  $\text{AWGN}(\sigma^2)$  in Table 2.1.

**Table 2.1.** Channel capacities and initial Bhattacharyya parameters of BSC, BEC and AWGN channels

Channel	Capacity	Initial Bhattacharyya parameter
BSC( $p$ )	$1 - H(p)$	$2\sqrt{p(1-p)}$
BEC( $\epsilon$ )	$1 - \epsilon$	$\epsilon$
AWGN( $\sigma^2$ )	$\frac{1}{2} \log_2 \left( 1 + \frac{E_s}{\sigma^2} \right)$	$e^{-\frac{E_s}{\sigma^2}}$

Polar codes are the codes that utilize the polarization effect. As discussed, the main idea is to send information on the channels whose capacities tend to 1 while freezing the bad channels whose capacities tend to 0. A polar code is defined with the parameter  $(N, K, \mathcal{A}, u_{\mathcal{A}^c})$  where  $N = 2^n$  is the codeword length,  $K$  is the number of information bits,  $\mathcal{A}$  is the information set, i.e. indices of the channels that information will be sent over, and  $u_{\mathcal{A}^c}$  being the vector of frozen bits. One can choose the frozen bits as desired. In this thesis, we will use all-zero vector for the choice of frozen bits. Since probability of block error for a  $(N, K, \mathcal{A}, u_{\mathcal{A}^c})$  polar code is upper-bounded by  $\sum_{i \in \mathcal{A}} Z(W_N^{(i)})$  [Arikan, 2012], the  $K$ -element information set  $\mathcal{A}$  is chosen from  $\{1, \dots, N\}$  such that the Bhattacharyya parameters satisfy  $Z(W_N^{(i)}) \leq Z(W_N^{(j)})$  for all  $i \in \mathcal{A}$  and  $j \in \mathcal{A}^c$ . In other words, we send the information bits from the channels with least Bhattacharyya parameters. The remaining  $N - K$  number of bits of the set  $\mathcal{A}^c$  are set to the frozen variables that the receiver knows the values and the channels that they are sent from. Furthermore, since  $Z(W)$  is channel specific as summarized in Table 2.1, polar codes are also specific to channels.

Recall the basic transform given in the beginning of this section, we have the encoding  $x_1 = u_1 + u_2$  and  $x_2 = u_2$ . We can represent this as  $X_1^2 = U_1^2 \cdot \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$  in vector

notation. Then, for the general case, given  $F \triangleq \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ , the encoded sequence  $x_1^N$  can be written as

$$x_1^N = u_1^N F^{\otimes n} \quad (2.26)$$

where  $F^{\otimes n}$  is the  $n^{\text{th}}$  Kronecker power of  $F$ , i.e.  $F^{\otimes n} = F^{\otimes n-1} \otimes F$ . In (Arıkan, 2009), Arıkan uses a permutation matrix  $B_N$  such that the encoding is  $x_1^N = u_1^N B_N F^{\otimes n}$ , however, he also adds that  $F^{\otimes n}$  can be used instead of  $B_N F^{\otimes n}$  to simplify the encoding, with the decoding done in bit-reversed index order. We will do the encoding as in (2.26) and change the order of decoding. The input sequence  $u_1^N$  is constructed as described in the previous paragraph.

To explain further, we will give a rate 0.5  $(N, K) = (8, 4)$  polar code explicitly. This code encodes  $K = 4$  bits of information while freezing  $N - K = 4$  bits. Which bits in the input sequence  $u_1^N$  are set to information or frozen depends on the Bhattacharyya parameters of the channels: The Bhattacharyya parameters of the channels  $W_N^{(i)}$ ,  $i = 1, \dots, 8$  where  $W = \text{BEC}(0.5)$  are shown in Figure 2.10. In this thesis, we assume the erasure probability of the channel is known beforehand, so that the polar code is designed (i.e. the Bhattacharyya parameters are calculated) according to the same erasure probability. Intuitively, since the code is specifically designed for the channel, the decoding performance would be better than the code performance with a constant design. Arıkan calls this as adaptive polar coding [Arıkan, 2008]. Choosing the 4 channels with the least Bhattacharyya parameters, the information bits are sent from  $(u_4, u_6, u_7, u_8)$  and the rest is set to frozen bits. As an example, if the information sequence is  $(1, 0, 1, 0)$ , the input sequence will be  $u_1^8 = (0, 0, 0, 1, 0, 0, 1, 0)$ . With the generator matrix  $F^{\otimes 3}$  given as

$$F^{\otimes 3} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

the encoded sequence is  $x_1^8 = (0,1,0,1,1,0,1,0)$ , with the overall complexity of  $O(N \log N)$ .

Once the encoded sequence is sent over the individual channels  $W$  and received as  $y_1^8$ , the decoding is done a bit-reversed order as mentioned: Representing  $y_1^8$  as  $y_0^7$ , i.e. letting the indices start from 0 and end at  $N - 1$ , the corresponding binary representations of the indices are (000,001,010,011,100,101,110,111). Reversing the bits, we get (000,100,010,110,001,101,011,111) which yields (0,4,2,6,1,5,3,7). So, instead of decoding in natural order (0,1,2,3,4,5,6,7) which means decoding  $y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8$ , decoding is done as in the aforementioned bit-reversed order resulting in decoding  $y_1, y_5, y_3, y_7, y_2, y_6, y_4, y_8$ .

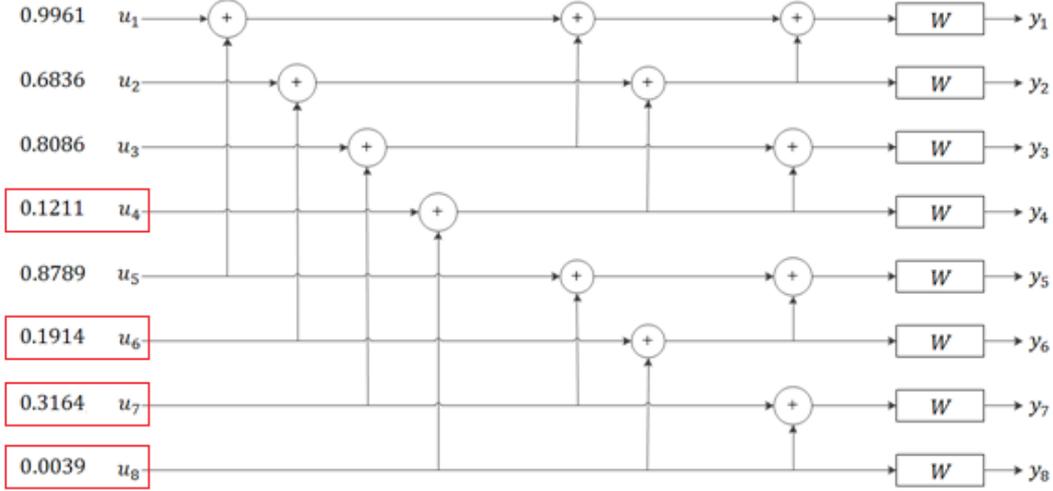
As discussed in Chapter 1, there are many suboptimum decoding algorithms, which can be used instead of the optimum ML decoding. In Arıkan's original scheme, successive cancellation (SC) decoding algorithm that uses the previously decoded bits without revising is used. For  $1 \leq i \leq N$ , the decoding algorithm is given as

$$\hat{u}_i = \begin{cases} 0, & \text{if } i \in \mathcal{A}^c \\ 1, & \text{if } i \in \mathcal{A} \end{cases} \quad (2.27)$$

$$d_i(y_1^N, \hat{u}_1^{i-1}) \triangleq \begin{cases} 0, & L_N^{(i)}(y_1^N, \hat{u}_1^{i-1}) \geq 0 \\ 1, & \text{otherwise} \end{cases}$$

where  $L_N^{(i)}$  is the log-likelihood ratio (LLR) given by

$$L_N^{(i)}(y_1^N, \hat{u}_1^{i-1}) \triangleq \log \frac{W_N^{(i)}(y_1^N, \hat{u}_1^{i-1}|0)}{W_N^{(i)}(y_1^N, \hat{u}_1^{i-1}|1)} \quad (2.28)$$



**Figure 2.10.** Bhattacharyya parameters of (8, 4) polar code designed for BEC(0.5) and selection of the indices (shown by red)

The decoding starts at the rightmost column (i.e. the channel level),  $L_1^{(1)}(y_i) = \frac{w(y_i|0)}{w(y_i|1)}$  which can be calculated directly, and continues to the left levels  $L_2^{(i)}, L_4^{(i)}$  and finally reaches the decision level  $L_N^{(i)}$ . For  $1 \leq j \leq N/2$ , the recursive formulas for  $L_N^{(i)}$  are found from (2.21) and (2.22) as

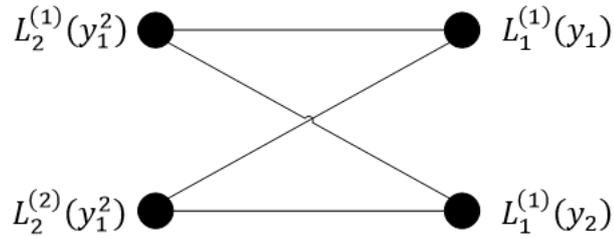
$$L_N^{(2j-1)}(y_1^N, \hat{u}_1^{2j-2}) = \frac{L_{N/2}^{(j)}(y_1^{N/2}, \hat{u}_{1,o}^{2j-2} \oplus \hat{u}_{1,e}^{2j-2}) L_{N/2}^{(j)}(y_{N/2+1}^N, \hat{u}_{1,e}^{2j-2}) + 1}{L_{N/2}^{(j)}(y_1^{N/2}, \hat{u}_{1,o}^{2j-2} \oplus \hat{u}_{1,e}^{2j-2}) L_{N/2}^{(j)}(y_{N/2+1}^N, \hat{u}_{1,e}^{2j-2})} \quad (2.29)$$

$$L_N^{(2j)}(y_1^N, \hat{u}_i^{2j-1}) \quad (2.30)$$

$$= \left[ L_{N/2}^{(j)}(y_1^{N/2}, \hat{u}_{1,o}^{2j-2} \oplus \hat{u}_{1,e}^{2j-2}) \right]^{1-2\hat{u}_{2j-1}} L_{N/2}^{(j)}(y_{N/2+1}^N, \hat{u}_{1,e}^{2j-2})$$

Similar to Fast-Fourier Transform structure, the SC decoding consists of butterfly structures. Equations (2.29) and (2.30) can be used to calculate upper-left and lower-left nodes' LLR values, respectively. Looking at the smallest butterfly as shown in Figure 2.11 and recalling from (2.9), the lower-left node uses upper's decoded bit to decode its own bit. Thus, in all butterflies, the lower-left node waits for the upper's decoding to result, which is completed using the results of the nodes on its right. Steps for an example decoding for  $(N, K, \mathcal{A}, u_{\mathcal{A}^c}) = (8, 5, \{3, 5, 6, 7, 8\}, (0, 0, 0))$  polar code are given explicitly in [Arıkan, 2009]. The complexity of this decoding scheme is also  $O(N \log N)$ .

Another decoding technique is called as the belief propagation (BP) decoding. Arıkan introduces this technique for polar codes [Arıkan, 2010b]. The structure of BP decoding is explained in the next section which is about the Reed Muller codes. It is shown that BP decoding of polar codes have a better BER performance over SC decoding [Akdođan, 2018; Chen, Niu, & Lin, 2012; Korada, 2009].



**Figure 2.11.** Successive cancellation (SC) decoder for polar code of code length 2.

## 2.4. Reed Muller Codes

As mentioned in Chapter 1 of this thesis, Reed Muller (RM) codes are one of the oldest, yet simplest codes in Coding Theory. They belong to the class of linear block codes over Galois field (GF(2)) with flexible parameters, thus draw attention due to ease of their decoding algorithms.

RM codes can be constructed in two ways: either inductively or using the Kronecker product. We start with the prior case, as given by Blahut [Blahut, 1984]. Denoting the element-wise (Hadamard) product of two vectors by  $\odot$ , we can write

$$\mathbf{A} \odot \mathbf{B} = (\mathbf{a}_0 \mathbf{b}_0, \mathbf{a}_1 \mathbf{b}_1, \dots, \mathbf{a}_{n-1} \mathbf{b}_{n-1})$$

where  $\mathbf{A} = \begin{bmatrix} \mathbf{a}_0 \\ \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_{n-1} \end{bmatrix}$  and  $\mathbf{B} = \begin{bmatrix} \mathbf{b}_0 \\ \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_{n-1} \end{bmatrix}$ , with  $\mathbf{a}_i$  and  $\mathbf{b}_i$  are single row vectors of the same

length. When constructing the generator matrix of the RM code, one simply takes all possible row products defined by  $\odot$ . Specifically, the generator matrix  $\mathbf{G}_{ind}$  of the  $r$ th-order RM code, constructed inductively, with codelength  $N = 2^n$  is given as

$$\mathbf{G}_{ind} = \begin{bmatrix} \mathbf{G}_0 \\ \mathbf{G}_1 \\ \vdots \\ \mathbf{G}_r \end{bmatrix}$$

where  $\mathbf{G}_0$  is the all-ones vector of length  $N$ ,  $\mathbf{G}_1$  is a  $n$ -by- $N$  matrix consisting of all binary  $n$ -tuples in its columns, and all other  $\mathbf{G}_m$ 's ( $2 \leq m \leq r$ ) are constructed from all possible  $m$ -row products from  $\mathbf{G}_1$  yielding  $\mathbf{G}_m$  to be an  $\binom{n}{m}$ -by- $n$  matrix. The minimum distance  $d^*$  of RM codes is  $2^{n-r}$  (see [Blahut, 1984] for proof) and since there are  $\sum_{i=0}^r \binom{n}{i}$  such rows, the number of information bits  $K$  in the  $(N, K)$  RM code is

$$K = \sum_{i=0}^r \binom{n}{i} = 1 + \binom{n}{1} + \dots + \binom{n}{r} \quad (2.31)$$

Note that RM codes can also be represented by  $\left\{ \begin{matrix} n \\ r \end{matrix} \right\}$  pair instead of  $(N, K)$ . However, due to notation used throughout the thesis, we will be using the  $(N, K)$  notation for the RM codes, as well.

As an example, if one wants to construct a  $(16, 11)$  RM code, then from (2.32) the order of this RM code is found to be  $r = 2$ . Then, the generator matrix of the  $(16, 11)$

RM code is  $\mathbf{G}_{ind} = \begin{bmatrix} \mathbf{G}_0 \\ \mathbf{G}_1 \\ \mathbf{G}_2 \end{bmatrix}$  with

- $\mathbf{G}_0$  is the all-ones vector:

$$\mathbf{G}_0 = [1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1] = [\mathbf{a}_0]$$

- $\mathbf{G}_1$  consists of all binary  $n = 4$ -tuples its columns:

$$\mathbf{G}_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \\ \mathbf{a}_4 \end{bmatrix}$$

- $\mathbf{G}_2$  is constructed by taking all  $m = 2$ -row products defined by  $\odot$ :

$$\mathbf{G}_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 \mathbf{a}_2 \\ \mathbf{a}_1 \mathbf{a}_3 \\ \mathbf{a}_1 \mathbf{a}_4 \\ \mathbf{a}_2 \mathbf{a}_3 \\ \mathbf{a}_2 \mathbf{a}_4 \\ \mathbf{a}_3 \mathbf{a}_4 \end{bmatrix}$$

Before mentioning the second construction method of RM codes, the rows of  $\mathbf{G}_{ind}$  can be reordered by row echelon permutations, which have no effect on the code. Once that is performed, it can be observed that  $\mathbf{G}_{ind}$  actually consists of  $F = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$  mentioned for the Polar code construction, which naturally leads to the Kronecker power construction of RM codes. For the construction of RM codes with Kronecker power, the  $n^{th}$  Kronecker power of  $F$  is taken, and  $K$  rows with minimum weight  $d^*$

are chosen. As an example, for the above (16, 11) RM code,  $d^* = 4$  and the rows of  $F^{\otimes 4}$  of weight at least 4 are chosen as the rows of the generator matrix  $\mathbf{G}_{Kron}$ :

$$\mathbf{G}_{Kron} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

It can be seen clearly that  $\mathbf{G}_{ind} = \mathbf{G}_{Kron}$  once the suitable row permutations are performed.

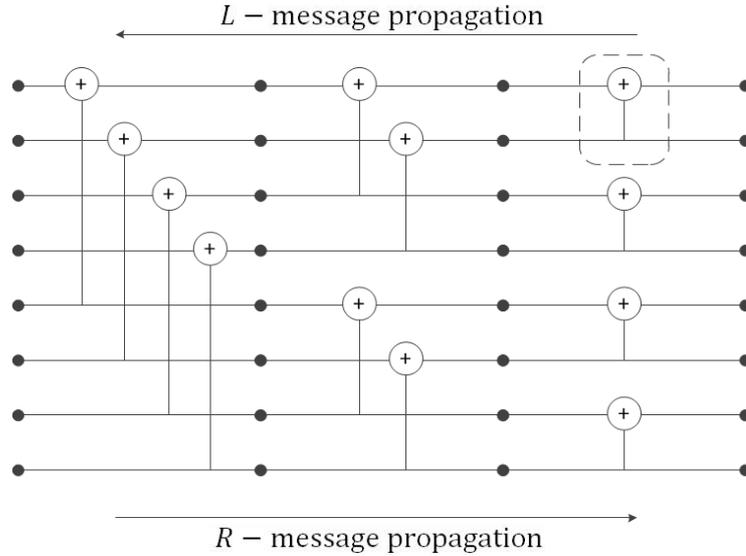
Similar to decoding of polar codes, a recursive decoding algorithm for RM codes was proposed [Dumer, 2017]. It is pointed out by Korada [Korada, 2009] that this algorithm is similar to the successive cancellation algorithm proposed by Arıkan [Arıkan, 2009]. RM codes are proven to be capacity achieving on both erasure and error channels [Abbe, Shpilka, & Wigderson, 2015; Kudekar et al., 2017] and outperform polar codes when MAP decoding is used [Mondelli, Hassani, & Urbanke, 2014], however they fall behind when SC decoding is used [Hashemi, Doan, Mondelli, & Gross, 2018]. Mondelli, Hassani and Urbanke thus use the two in a hybrid structure and benefit from both under certain channel conditions [Mondelli et al., 2014]. Forney proposes another decoding called ‘‘Belief Propagation’’ (BP) decoding for codes on graphs [Forney, 2001], among which RM codes also belong. BP decoding performance lies nearly halfway between that of MAP and SC decoding, and in fact, SC decoder is a particular instance of BP decoder [Korada, 2009]. BP decoding can be used for both RM and polar codes, but it is shown that performance of polar codes under BP decoding is better than that of RM codes [Arıkan, 2010a].

BP decoding is a message passing algorithm on factor graphs, and it is based on Gallager’s LDPC decoding [Elkelesh, Cammerer, Ebada, & Ten Brink, 2017]. It

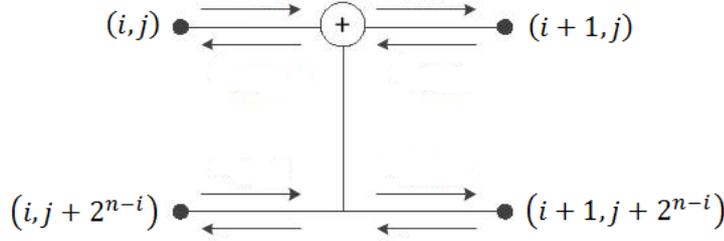
consists of  $n = \log_2 N$  stages and  $\frac{N}{2} \log_2(N)$  processing elements (PE) depicted with dashed box in Figure 2.12 and detailed in Fig.1b. Figure 2.12 is also called as “factor graph”. Denoting the nodes of PE with  $(i, j)$  where  $i$  is the stage ( $1 \leq i \leq n + 1$ ) and  $j$  is the bit index ( $1 \leq j \leq N$ ), at each PE, messages are passed from left to right (denoted as  $L_{i,j}^{(t)}$ ) and from right to left (denoted as  $R_{i,j}^{(t)}$ ) by

$$\begin{aligned}
 L_{i,j}^{(t+1)} &= f\left(L_{i+1,j}^{(t)}, L_{i+1,j+N_i}^{(t)} + R_{i,j+N_i}^{(t)}\right) \\
 L_{i,j+N_i}^{(t+1)} &= L_{i+1,j+N_i}^{(t)} + f\left(L_{i+1,j}^{(t)}, R_{i,j}^{(t)}\right) \\
 R_{i+1,j}^{(t+1)} &= f\left(R_{i,j}^{(t)}, L_{i+1,j+N_i}^{(t)} + R_{i+1,j+N_i}^{(t)}\right) \\
 R_{i+1,j+N_i}^{(t+1)} &= R_{i,j+N_i}^{(t)} + f\left(R_{i,j}^{(t)}, L_{i+1,j}^{(t)}\right)
 \end{aligned} \tag{2.32}$$

where  $N_i = 2^{n-i}$ ,  $t = 0, 1, \dots$  is the time index and  $f(a, b) = \log \frac{1+e^{a+b}}{e^a+e^b}$ .



**Figure 2.12.** Belief propagation (BP) decoder construction for RM codes of code length 8. The dashed line shows a processing element (PE)



**Figure 2.13.** Details of a processing element (PE) used in a BP decoder

The initial leftmost  $R_{n+1,j}^{(0)}$  messages, denoting the *a priori* information at the decoder, and the rightmost  $L_{n+1,j}^{(0)}$  messages, denoting the channel output, are initialized as

$$L_{n+1,j}^{(0)} = \log \frac{P(x_j = 0|y_j)}{P(x_j = 1|y_j)} \quad (2.33)$$

$$R_{n+1,j}^{(0)} = \begin{cases} 0, & \text{if } j \text{ is information index} \\ \infty, & \text{if } j \text{ is frozen index} \end{cases}$$

while all other nodes at time  $t = 0$  are set to 0. The decoding iteratively continues until a predefined number of iterations  $N_{iter}$  are performed. Assuming the decoding is finished at  $t = T$ , the LLRs of the estimates  $\hat{\mathbf{u}}$  of the input  $\mathbf{u}$  and the transmitted codeword  $\hat{\mathbf{x}}$  are calculated as

$$\hat{u}_j = \begin{cases} 0, & \text{if } L_{1,j}^{(T)} + R_{1,j}^{(T)} \geq 0 \\ 1, & \text{otherwise} \end{cases} \quad (2.34)$$

$$\hat{x}_j = L_{n+1,j}^{(T)} + R_{n+1,j}^{(T)}$$

Since  $N_{iter}$  may result in decoding latency, the iterations may also be stopped when early stopping conditions are met. Although suggested primarily for the polar codes, we prefer to mention these in this section due to the given context of BP decoding, yet without giving too much detail. Elkelesh, Ebada, Cammerer and Ten Brink describe 3 stopping criteria called as the practical, perfect knowledge-based, and CRC-aided stopping conditions [Elkelesh, Ebada, Cammerer, & Ten Brink, 2018a]. The first one is simply the generator matrix-based ( $\mathbf{G}$  – based) condition where the iterations stop

when  $\hat{\mathbf{x}} = \hat{\mathbf{u}}\mathbf{G}$  is satisfied and thus  $\hat{\mathbf{u}}$  is a valid estimate. The second one continues iterating until  $\hat{\mathbf{u}} = \mathbf{u}$  is satisfied, which requires the knowledge of the information bit and thus may be called as “genie-aided”, and the last one requires an outer high rate cyclic redundancy check (CRC) code that checks upon the information bits. Yuan and Parhi propose a minimum LLR-based criterion that stops the iteration when all LLRs of  $\hat{\mathbf{u}}$  exceed a given threshold value [Yuan & Parhi, 2014]. Furthermore, an adaptive stopping condition that is based on channel condition estimation that determines this threshold with respect to the channel SNR is proposed. In summary, both of the aforementioned proposals improve the performance of BP decoders.

## 2.5. The Wireless Fading Channel

As mentioned before, propagation of signals in wireless channel is by electromagnetic waves. Assuming no phase offset, we can denote the band pass input signal as  $s(t) = \Re\{u(t)e^{j2\pi f_c t}\}$  where  $u(t)$  is the complex envelope of  $s(t)$  and  $f_c$  is the carrier frequency. Due to reflection, diffraction and scattering, there are many paths of a single transmitted signal, each with different delays and Doppler phase shifts. Then, the resulting received signal is sum of all delayed and attenuated replicas of the transmitted signal:

$$r(t) = \Re \left\{ \sum_{n=0}^{N(t)} \alpha_n(t) u(t - \tau_n(t)) e^{j2\pi[f_c(t - \tau_n(t)) + \phi_{D_n}]} \right\} \quad (2.35)$$

where  $n$  denotes the  $n$ th path ( $n = 0$  denotes the line-of-sight (LOS) path);  $\alpha_n(t)$  being the attenuation factor (which is a function of path loss),  $\tau_n(t)$  being the time delay and  $\phi_{D_n}$  being the Doppler phase shift of  $n$ th path (Note that the Doppler shift on  $n$ th path is  $f_{D_n}(t) = -f_c \tau_n(t)$  and  $\phi_{D_n} = \int_t 2\pi f_{D_n}(t) dt$ ). Letting  $\phi_n(t) = 2\pi f_c \tau_n(t) - \phi_{D_n}$  so that it represents both delay and Doppler phase shift, the simplified version of received signal is

$$r(t) = \Re \left\{ \left[ \sum_{n=0}^{N(t)} \alpha_n(t) e^{-j\phi_n(t)} u(t - \tau_n(t)) \right] e^{j2\pi f_c t} \right\} \quad (2.36)$$

On the other hand, if we model the channel as a linear time varying system with a baseband channel impulse response  $h(t, \tau)$ , we can simply write  $r(t) = \Re \{ [\int_{-\infty}^{\infty} h(t, \tau) u(t - \tau) d\tau] e^{j2\pi f_c t} \}$ . Comparing this to (2.37), we have the impulse response for a fading channel as

$$h(t, \tau) = \sum_{n=0}^{N(t)} \alpha_n(t) e^{-j\phi_n(t)} \delta(\tau - \tau_n(t)) \quad (2.37)$$

where  $\delta(\cdot)$  is the dirac delta function. For the discrete case, we can naturally write  $y[n] = \sum_l h_l[m] x[m - l]$  such that  $h_l[m]$  is the  $l$ th channel filter tap at time  $m$  [Rappaport, 1996]. We will not go into further detail as it is sufficient to keep in mind that the channel impulse response is a function of attenuation, delay and phase shifts for the scope of this thesis.

As mentioned before, fading is variation in signal power over time and frequency, and can be divided into two: large-scale and small-scale propagation effects. The former effect is mainly due to path loss and blocking of objects, i.e. shadowing, in relatively large distances and time durations, while the latter effect characterizes the rapid signal power fluctuations that happen in much shorter distances and time. Both propagation effects are modeled statistically. Without loss of generality, the term fading alone is used for small-scale propagation effects and we will be concerned small-scale fading in the simulations performed in Chapter 3.

Factors affecting small-scale fading can be stated as multipath propagation, speed of the mobile station with respect to base station or the surrounding objects, and the transmission bandwidth of the signal [Rappaport, 1996]. From (2.35), due to randomness and dependency on delay and Doppler shift, it is evident that there would some key parameters when defining a fading channel such as how fast the channel

changes over time or frequency. *Coherence time* represents how fast the channel is changing over time and is a function of Doppler shifts of different paths contributing to the same filter tap  $h_l[m]$ : If we denote *Doppler spread* as maximum difference of Doppler shifts, i.e.  $D_s := \max_{i,j} f_c |\tau_i(t) - \tau_j(t)|$ , then the coherence time is  $T_c = 1/4D_s$ . When the codeword length spans multiple channel fades, or equivalently many coherence periods  $T_c$ , we have the *fast fading channel*. In fast fading channel, the channel changes are fast that the bit errors for the individual bits are independent [Lee, 2004]. Going further, if channel tap  $h_l[m]$  remains constant over  $T_c$  symbols, we have the simple *block fading channel* that can be thought as parallel channels with filter taps being constant within the block, while being i.i.d. among different blocks. We assume the block fading channel model throughout the simulations in Chapter 3.

## 2.6. Further Notes and Literature Review

In this section, we give some further notes and try to summarize the existing work related to our work presented in this thesis.

### 2.6.1. Decoders for RM and Polar Codes

From the point of view of Kronecker product construction of RM codes, it is obvious that RM and polar codes are very similar, in terms of their generator matrices. The difference is that RM codes use minimum distance rule while the polar codes use minimum Bhattacharyya parameter-rule. In other words, polar codes are channel-specific whereas the RM codes are channel-independent. It is shown that for  $n \leq 4$ , the generator matrices of the two codes are exactly the same, but after  $N = 32$ , the rows of the generator matrices start to differ [Arıkan, 2008]. Nonetheless, they can be decoded using the same decoders. As a result, due to the similarities in both encoding and decoding, they remain to be close relatives and they are often compared in terms of error rate performance. While doing so, further decoding techniques that can be used in decoding either of the two codes are used:

*i. CRC-aided Decoding*

CRC code is an error-detecting code that is based on *cyclic* codes and *checks* the data by adding some *redundancy*. It is commonly used in various areas such as telecommunication standards such as 3GPP or GSM, mobile networks or computer architectures. They are denoted as CRC- $n$  where  $n$  is the number of redundancy bits. Since all of the decoding techniques may well be improved with the help of a CRC code, we will not be going much further in detail here (and also in the following described decoding techniques), except mentioning that it was shown that a SCL decoding of polar codes with CRC-16 outperformed state-of-the-art LDPC codes [Tal & Vardy, 2015]. For other applications of CRC codes used in decoding of polar codes, preferred readers may read [B. Li, Shen, & Tse, 2014; Murata & Ochiai, 2017; Niu & Chen, 2012].

*ii. Permuted Decoding*

Due to iterative construction of polar codes, one can easily manipulate the stages of the factor graph, such that the stages are permuted. It was Hussami, Korada and Urbanke [Hussami et al., 2009] and Korada alone [Korada, 2009] to first mention permutation of layers of the factor graph for decoding of the polar codes, leaving it as an open problem. Since there are  $n = \log_2 N$  layers in a factor graph, there are  $n!$  ways to construct it, irrespective of the type of decoder used. As a result, we will mention the permuted SC and BP decoding (PSCD and PBPD, respectively) here.

Vangala, Viterbo and Hong give the proof of encoder's permutation-invariance and uses the permutation over the SC decoder. They show that although the performance is degraded when PSCD is used for the polar codes designed for the standard SC-decoder (which is the bit-reversed decoding order as given in the original construction of polar codes), the PSCD performance is exactly the same as SCD when the polar code construction is matched to the permutation used at the decoder (i.e., the order of Bhattacharyya parameters of the bit channels are in the same order as the decoding order) They also conclude that using the latter case, if exactly the opposite order of the layers (with respect to Arıkan's original construction) is used with a matched

construction, the decoding order becomes the natural order which results in less complexity and latency of the decoder [Vangala, Viterbo, & Hong, 2014].

Permutation on the BP decoder is also used [Akdoğan, 2018; Doan, Hashemi, Mondelli, & Gross, 2018; Elkelesh et al., 2018a]. Elkelesh, Ebada, Cammerer and Ten Brink use these different permutations in series such that if one permuted factor graph fails to decode the received codeword, a permuted version of it is used until a predefined number of maximum permutations is reached. They conclude that using more than one representation (i.e., permutation) of the factor graph results in improved decoding performance, and even outperforms SCL decoder when a proper stopping criterion (for example, perfect knowledge-based one) is used [Elkelesh et al., 2018a]. Unlike the random choice of permutation used in the previously mentioned work, Doan, Hashemi, Mondelli and Gross propose a method to construct a predetermined set of permutations, which consists of only the good permutations of the original factor graph, and it is shown to improve FER performance. Interestingly, they also mention that the good permutations are those which are obtained by permuting the leftmost side of the factor graph [Doan et al., 2018] (Although the authors conclude it as “rightmost” instead of “leftmost”, the factor graph they take as reference is the opposite of the one used in this thesis, as shown in Figure 2.12. Thus, we give their result the opposite way, without loss of information). This conclusion is also verified by numerical calculations by Akdoğan, who also investigates the use of multiple factor graphs, both dependently (by letting the newly constructed factor graph’s erasures to be filled by that of previous factor graph) and independently (by sending the undecoded codewords to the next factor graph), and concludes that dependent use of multiple factor graphs outperform the independent use under BP decoding [Akdoğan, 2018].

Another way of using the factor graph permutations both for the polar and RM codes, referred to as Successive Permutation SC List Decoding (SPSCL), is discussed by Hashemi, Doan, Mondelli and Gross. Using the code partitioning described by Hassani *et al.* [Hassani et al., 2018], it is pointed out that there are actually more than

$n!$  permutations, and thus they propose a method that constructs the best permutation “on the fly”, i.e. during the course of decoding, that picks the permutation with the most reliable LLR values. Once conducted both for the SC and SCL decoders, this scheme shows improved FER performance in decoding both of the codes [Hashemi et al., 2018].

### iii. List Decoding

List decoding was actually mentioned for RM codes using their recursive encoding structure called as the *Plotkin* construction [Dumer & Shabunov, 2017]. Arıkan also mentions this construction for Polar codes [Arıkan, 2008] and [Arıkan, 2010a]. Plotkin construction basically decomposes the  $\begin{Bmatrix} n \\ r \end{Bmatrix}$  code onto subblocks  $\begin{Bmatrix} n-1 \\ r \end{Bmatrix}$  and  $\begin{Bmatrix} n-1 \\ r-1 \end{Bmatrix}$  until the repetition codes  $\begin{Bmatrix} g \\ 0 \end{Bmatrix}$  for any  $g = 1, \dots, n-r$  and full spaces  $\begin{Bmatrix} h \\ h \end{Bmatrix}$  for any  $h = 1, \dots, r$  are reached. As a result, tree-like construction is formed and recursive encoding and decoding can be performed (for details, see [Dumer, 2004, 2006]).

List decoding of polar codes are considered both for the SC and BP decoders. Tal and Vardy propose list decoding to the SC decoder, referring to as SCL decoding with a parameter  $L$  called the list size [Tal & Vardy, 2015]. The decoding is performed successively one-by-one as in the original SC decoder, but the SCL decoder takes  $L$  decoding paths into account when decoding the next information bit: When decoding an unfrozen bit  $\hat{u}_{i+1}$ , the decoder splits the previous  $L$  decoding paths (used for the decoding of the previous information bits  $\hat{\mathbf{u}}_0^i$ ) into two such that  $\hat{u}_{i+1}$  can either be 0 or 1, and then keeps the most likely  $L$  paths for the next decoding phase. At the end, the decoder chooses the most likely path among  $L$  paths, giving a single decoded codeword. When  $L = 1$ , the SCL decoder is simply the original SC decoder where it decides a decoded information bit to be 0 or 1 instantly at each decoding phase (without keeping track of decoding paths). It is observed that error rate performance is improved with increasing list size, with a cost of increased complexity and latency. Hardware improvements regarding this issue, such as partitioning the decoding paths

(referred to as partitioned-SCL (PSCL) to reduce memory requirements exist [Hashemi, Balatsoukas-Stimming, Giard, Thibeault, & Gross, 2016], but hardware implementation considerations are not within scope of this thesis. Nonetheless, the performance is comparable to state-of-the-art LDPC and Turbo codes, and indeed SCL decoding of Polar codes outperforms LDPC codes used in WiMAX when CRC precoding is used [Tal & Vardy, 2015].

Unlike the serial use of permuted factor graphs discussed in PBP decoders, Elkelesh, Ebada, Cammerer and Ten Brink use the BP decoders in parallel, resulting in the BP-List (BPL) decoding of polar codes. The proposed method is to select  $n - 1$  cyclic shifts among different factor graph representations, decode the received codeword using  $L$  parallel BP decoders, and finally picking the codeword closest (in terms of Euclidean distance) to the received codeword. They claim that this proposed scheme is the best iterative decoder, in terms of soft-decoding and low latency [Elkelesh, Ebada, Cammerer, & Ten Brink, 2018b]. They also use RM and polar codes together, which is described in Chapter 4 of this thesis.

## **2.6.2. Fading Channels and Polar Codes**

Existing work relating to polar codes and fading channels do not date far back. Our underlying motivation to investigate this relationship is mainly due to the work of Liu, Hong and Viterbo [Liu, Hong, & Viterbo, 2017]. As the probability of error in a fading channel is inversely proportional to the channel gain (or fading coefficient), they claim that the fading channels are naturally polarized, which they refer to as “fading polarization”. They propose a new method to calculate the Bhattacharyya parameters specifically for the fading channel thus match the polar code construction to the fading channels. This method is found to provide 1.5 dB gain over LDPC codes at block error rate  $10^{-4}$ .

Bravos-Santos and Trifonov both narrow the fading channels to only the Rayleigh fading and consider the polar codes for such channels [Bravo-Santos, 2013; Trifonov,

2015]. Bravo-Santos constructs polar codes assuming the channel statistics such as either channel distribution or channel state information; i.e. CDI or CSI, is known for binary input- and for block Rayleigh fading channels. Compared to Turbo and LDPC codes, it is shown that the suggested scheme is closer to the theoretical limit when large codelengths are used [Bravo-Santos, 2013]. Similar to fading polarization mentioned by Liu *et al.* [S. Liu et al., 2017], Trifonov models the polarized subchannels as fading channels whose gains have Chi ( $\chi$ ) distribution which is a general case of Rayleigh distribution. It is noted that this modelling can be used to estimate the error probabilities in the polarized subchannels. Furthermore, it is pointed out that classical polar codes perform poorly in the fading channels unless the code is optimized for the Rayleigh channel, frozen bits are dynamically set to linear combinations of other symbols, and sequential or list decoding is used. For the latter case, it is shown that use of dynamic frozen symbols provides significant gain over similar LDPC code [Trifonov, 2015].

Continuing with the block fading channels, Boutros and Biglieri state that unlike the natural construction of polar codes which is deterministic; in block fading channels, the channel polarization can be thought as multiple parallel channels having different mutual information. Within a block of  $N$  symbols, they assume only two fading states with no specific distribution and affecting  $N/2$  symbols irrespective of the order, which they refer to as “multiplexing”. As a result, they observe the polarization effect on 3 different such multiplexed structures by analyzing the mutual information outage probabilities [Boutros & Biglieri, 2013].

Similar to the work of Boutros and Biglieri, Si, Köylüoğlu and Vishwanath propose a hierarchical polar coding scheme for block fading binary symmetric and additive exponential noise channels without CSI at the transmitter. For the block fading BSC case, they start with two channel states that one block can be in, and generalize to  $S$  states [Si, Köylüoğlu, & Vishwanath, 2014]. Observing that this work is mainly theoretical, we leave the details of it to Chapter 3 where we try to investigate the bit error rate performance of this proposed scheme.

## CHAPTER 3

### HIERARCHICAL POLAR CODING SCHEME FOR FADING BINARY SYMMETRIC CHANNELS

In this chapter, we compare the hierarchical polar coding scheme for binary fading channels proposed by Si, Köylüoğlu and Vishwanath in 2014 [Si et al., 2014] with Arıkan’s original polar code scheme [Arıkan, 2009]. We first briefly explain the scheme and then compare the BER (bit error ratio) performance of the hierarchical polar coding scheme in [Si et al., 2014] with that of the plain polar coding [Arıkan, 2009] under similar conditions.

#### 3.1. Proposed Scheme by Si, Köylüoğlu and Vishwanath

Si, Köylüoğlu and Vishwanath propose a hierarchical coding scheme that uses multiple polar coding phases for block fading channels with additive binary and exponential noise channels. In this thesis, we call this code the “SKV-code”; and we focus on the fading BSC case as the model of the AWGN block fading channel with BPSK modulation. The authors assume that only the decoder knows the channel state information (in short we will denote this as CSI-D as “channel state information at the decoder”), while the transmitter only knows the channel state statistics. For simplicity, we focus on the case of two-state fading channel, where there are only two fading states that a block may encounter, although in [Si et al., 2014], generalization to  $S$  states is also given.

Referring to Figure 2.3(b), the output  $Y$  of a fading BSC with input  $X$  can be represented as

$$Y_{b,i} = X_{b,i} \oplus Z_{b,i}, \quad i = 1, \dots, N, \quad b = 1, \dots, B, \quad (3.1)$$

where  $N$  is the block length,  $B$  is the number of blocks, and  $Z_{b,i}$ 's are identically Bernoulli distributed within a block and independent over fading blocks. As mentioned before, we assume only two states; that is, with probability  $q_1$ , the block  $b$  can be in State 1 and with probability  $q_2 = 1 - q_1$ , it can be in State 2. For the blocks in State 1, the noise  $Z_{b,i}$  is Bernoulli distributed with parameter  $p_1$ . Likewise, for the blocks in State 2, the noise  $Z_{b,i}$  is Bernoulli distributed with parameter  $p_2 \leq p_1$ . These probabilities will be used to model the AWGN channel with BPSK modulation as discussed below.

As mentioned in Section 2.5, when we talk about fading channels, for the discrete AWGN case, we have  $Y = hX + Z$ , where  $h$  is the channel gain and  $Z$  is the Gaussian noise. With BPSK modulation, this channel can be considered as a binary input and binary output channel with transition probability relating to the AWGN channel state. In other words, since the channel is assumed to be constant over a block, the channel gain  $h_{b,i} = h_s, \forall i = 1, \dots, N$ , with probability  $q_s$  for  $s \in \{1,2\}$  and the equivalent fading BSCs have the crossover probabilities

$$p_s \triangleq P\{Z_{b,i} = 1\} = 1 - \Phi(h_s \sqrt{SNR}), \quad s \in \{1,2\} \quad (3.2)$$

where  $\Phi(x)$  is the cumulative normal function of the Gaussian distribution (See Appendix A for proof). As a result, the two fading blocks are equivalent to two BSCs:  $W_1 = \text{BSC}(p_1)$  and  $W_2 = \text{BSC}(p_2)$ .

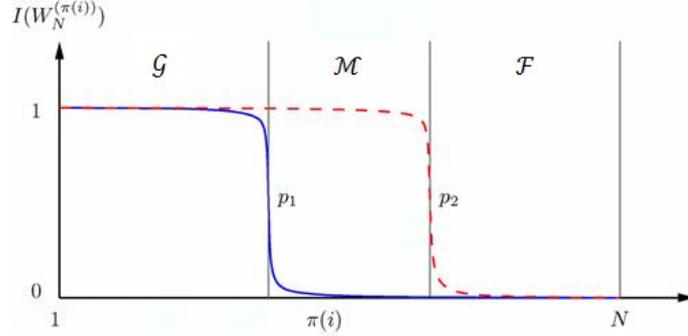
The overall capacity for the 2-state fading BSC is

$$\begin{aligned} C &= \sum_{s=1}^2 q_s [1 - H(p_s)] \\ &= q_1 [1 - H(p_1)] + (1 - q_1) [1 - H(p_2)] \end{aligned} \quad (3.3)$$

and the authors state that the proposed polar coding scheme achieves this capacity without channel state information at the transmitter [Si et al., 2014].

It is assumed that  $0.5 \geq p_1 \geq p_2$ , so we will call  $W_1$  as the bad (or degraded) channel and  $W_2$  as the good (or superior) channel. Since Bhattacharyya parameters for the construction of polar codes are required, the information bit indices of the bad channel are a subset of that of the good channel. This is given in Lemma 1 of [Si et al., 2014]. After reordering the indices, they are grouped into three sets as shown in Fig.3.1.

1. Set  $\mathcal{G}$  where both channels' capacities tend to 1 (i.e. both are good channels).
2. Set  $\mathcal{M}$  where the degraded channel's capacity tends to 0, while that of the superior tends to 1.
3. Set  $\mathcal{F}$  where both channels' capacities tend to 0.



**Figure 3.1.** Choice of information bits in a codeword of length  $N$  according to the encoding scheme proposed in [Si et al., 2014] for two fading states (reproduced from [Si et al., 2014]).

Then, if we denote the sets of information bits as  $\mathcal{A}_1$  and  $\mathcal{A}_2$  for the degraded and superior channels, respectively, we can write  $\mathcal{A}_1 = \mathcal{G}$  and  $\mathcal{A}_2 = \mathcal{G} \cup \mathcal{M}$ , and the following relations are formed considering the sizes of these sets:

$$\begin{aligned}
 |\mathcal{G}| &= |\mathcal{A}_1| = [1 - H(p_1) - \epsilon]N \\
 |\mathcal{M}| &= |\mathcal{A}_2| - |\mathcal{A}_1| = [H(p_1) - H(p_2)]N \\
 |\mathcal{F}| &= N - |\mathcal{A}_2| = [H(p_2) + \epsilon]N
 \end{aligned} \tag{3.4}$$

where  $H(\cdot)$  is the binary entropy function and the offset  $\epsilon$  in  $[0, 1]$  approaches 0 as  $N \rightarrow \infty$ . One should notice that, the overall code rate is  $R = q_1 R_1 + (1 - q_1) R_2$ , where  $R_1 = |\mathcal{G}|/N = [1 - H(p_1) - \epsilon]$  and  $R_2 = [|\mathcal{G}| + |\mathcal{M}|]/N = [1 - H(p_2) - \epsilon]$ . So, the code rate  $R_1$  is less than the 2-state fading channel capacity  $C|_{q_1=1} = 1 - H(p_1)$  given by (3.3) and  $R_2$  is less than  $C|_{q_1=0} = 1 - H(p_2)$  only for nonzero values of the offset  $\epsilon$ .

Looking at Fig.3.1, regardless of the fading state, the channel with indices in the set  $\mathcal{G}$  always polarizes to a good channel (the capacity tends to 1) and in the similar manner, the channel with indices in the set  $\mathcal{F}$  always polarizes to a bad channel (the capacity tends to 0). Then, we can send the information bits at indices belonging to set  $\mathcal{G}$  reliably, while we can “freeze” the indices belonging to the set  $\mathcal{F}$ . However, the indices over the set  $\mathcal{M}$  behave differently: With probability  $q_1$ , the constructed channel is in the degraded state, while with probability  $1 - q_1$ , the constructed channel is in the superior state. In other words, the information bits are sent unreliably with probability  $q_1$ . So, the channels in set  $\mathcal{M}$  can be modeled as a BEC with erasure probability of  $q_1$ :  $\tilde{W} \triangleq BEC(q_1)$ . As a result, the uncertainty in layer  $\mathcal{M}$  can be overcome by exploiting an overlaid BEC over the fading BSCs.

With this being said, the encoder hierarchically uses two phases to construct a codeword of length  $NB$  as explained in the next section, followed by the block fading channel modelling and finally the three-phased decoder to decode the information bits.

### 3.1.1. Hierarchical Encoding

#### 3.1.1.1. Phase 1 – BEC Encoding

In this phase,  $|\mathcal{M}|$  polar codewords specified by the parameter  $(B, |\tilde{\mathcal{A}}|, \tilde{\mathcal{A}}, 0)$  are constructed by setting the frozen variables to 0, where  $\tilde{\mathcal{A}}$  is the information set for  $\tilde{W} = BEC(q_1)$  such that

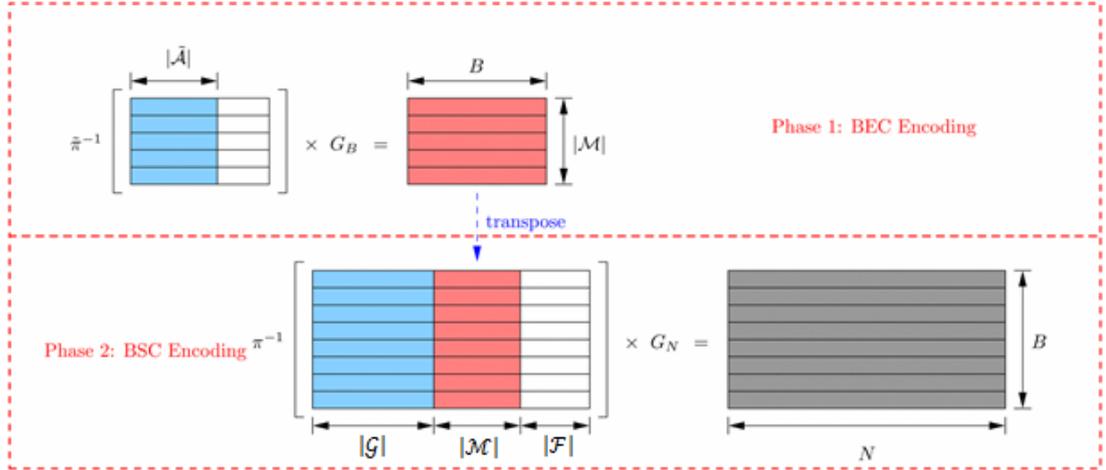
$$|\tilde{\mathcal{A}}| = (1 - q_1 - \delta)B \quad (3.5)$$

where  $0 \leq \delta \ll 1 - q_1$  and vanishes as  $B$  gets large enough. The encoder generates  $|\mathcal{M}| \times B$  bits denoted as  $\tilde{\mathbf{U}}$  such that

$$\tilde{\mathbf{U}} = \begin{bmatrix} u_{11} & u_{12} & u_{13} & \dots & \dots & u_{1B} \\ u_{21} & u_{22} & u_{23} & \dots & \dots & u_{2B} \\ \vdots & \dots & \dots & \dots & \dots & \vdots \\ u_{|\mathcal{M}|1} & u_{|\mathcal{M}|2} & u_{|\mathcal{M}|3} & \dots & \dots & u_{|\mathcal{M}|B} \end{bmatrix} = \begin{bmatrix} \mathbf{U}_1 \\ \mathbf{U}_2 \\ \vdots \\ \mathbf{U}_{|\mathcal{M}|} \end{bmatrix}$$

### 3.1.1.2. Phase 2 – BSC Encoding

In this phase,  $B$  polar codewords specified by the parameter  $(N, |\mathcal{G}|, \mathcal{G} \cup \mathcal{M}, 0)$ , where  $\mathcal{G}$  is the information set for  $W_1 = BSC(p_1)$ , are produced. Unlike setting all of the frozen bits to 0, the output of the first phase is transposed and used as frozen bits, along with the remaining zeros. The encoder thus encodes  $\pi^{-1}([\mathbf{U} \mid \tilde{\mathbf{U}}^T \mid \mathbf{0}])$  (where  $\pi(\cdot)$  is the permutation to order the Bhattacharyya parameters in ascending manner, for the degraded channel) and generates an overall codeword of length  $NB$ . The overall encoding scheme is shown in Fig.3.2.



**Figure 3.2.** Polar encoder proposed in [Si et al., 2014] for a fading binary symmetric channel with two states (reproduced from [Si et al., 2014])

### 3.1.2. The Block Fading Channel

The output of the encoder denoted as  $\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & \dots & x_{1N} \\ x_{21} & x_{22} & x_{23} & \dots & \dots & x_{2N} \\ \vdots & \dots & \dots & \dots & \dots & \vdots \\ x_{B1} & x_{B2} & x_{B3} & \dots & \dots & x_{BN} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_B \end{bmatrix}$

such that  $\mathbf{X} = X_{b,i}^N$ ,  $b = 1:B$ ,  $i = 1:N$  is transmitted from the fading BSCs as follows:

$$Y_{b,i} = X_{b,i} \oplus Z_{b,i} \text{ such that } \begin{cases} Z_{b,i} \sim \text{Bern}(p_1), \text{ channel state}_b = \text{"bad"} \\ Z_{b,i} \sim \text{Bern}(p_2), \text{ channel state}_b = \text{"good"} \end{cases} \quad (3.6)$$

In other words, if the block  $b$  is in the degraded (“bad”) channel state (which happens with probability  $q_1$ ), the noise added to this block is distributed as Bernoulli random variable with probability  $p_1$ , and if the block is in the superior (“good”) channel state (with probability  $q_2 = 1 - q_1$ ), the noise added to this block is distributed as Bernoulli random variable with probability  $p_2$ . After  $\mathbf{X}$  is transmitted from the fading channel, the received codeword  $\mathbf{Y}$  such that  $\mathbf{Y} = Y_{b,i}^N$ ,  $i = 1:N$ ,  $b = 1:B$  is decoded using three phases, with the channel state information being known at the receiver as shown in Figure 3.3.

### 3.1.3. Hierarchical Decoding

#### 3.1.3.1. Phase 1 – BSC Decoding 1

In the first phase of decoding, the decoder uses the classical BSC( $p_2$ )-SC (successive cancellation) polar decoder with parameter  $p_2$ :

$$\hat{u}_i^{(b)} \triangleq \begin{cases} 0 & \text{if } i \in \mathcal{F} \\ d_{2,i}(y_{1:N}^{(b)}, \hat{u}_{1:i-1}^{(b)}) & \text{if } i \in \mathcal{G} \cup \mathcal{M}, \end{cases} \quad i = 1:N$$

$$d_{2,i}(y_{1:N}^{(b)}, \hat{u}_{1:i-1}^{(b)}) \triangleq \begin{cases} 0, & \text{if } \frac{W_{2,N}^{(i)}(y_{1:N}^{(b)}, \hat{u}_{1:i-1}^{(b)} | 0)}{W_{2,N}^{(i)}(y_{1:N}^{(b)}, \hat{u}_{1:i-1}^{(b)} | 1)} \geq 1 \\ 1, & \text{otherwise.} \end{cases} \quad (3.7)$$

This procedure decodes the information bits in blocks with respect to the superior channel states reliably; however, the ones with respect to the degraded channel states cannot be decoded reliably, because the frozen bits of the degraded channel states ( $\mathcal{M} \cup \mathcal{B}$ ) are unknown due to the unknown set  $\mathcal{M}$ . Thus, a  $B \times |\mathcal{M}|$  matrix  $\widehat{\mathbf{U}}$  is constructed by choosing the rows corresponding to the superior state directly from this phase's output and the rows corresponding to the degraded state are set to erasures, which are then sent to the next phase.

### 3.1.3.2. Phase 2 – BEC Decoding

In this phase, the frozen bits with respect to the degraded channel states are decoded using the classical BEC( $q_1$ )-SC decoder with parameter  $q_1$ :

$$\hat{v}_j^{(k)} \triangleq \begin{cases} 0 & \text{if } i \in \tilde{\mathcal{A}}^c \\ \tilde{d}_j(\tilde{u}_{1:|\mathcal{M}|}^{(k)}, \hat{v}_{1:j-1}^{(k)}) & \text{if } i \in \mathcal{A} \end{cases}, \quad j = 1:B$$

$$\tilde{d}_j(\tilde{u}_{1:|\mathcal{M}|}^{(k)}, \hat{v}_{1:j-1}^{(k)}) \triangleq \begin{cases} 0, & \text{if } \frac{\tilde{W}_N^{(j)}(\tilde{u}_{1:|\mathcal{M}|}^{(k)}, \hat{v}_{1:j-1}^{(k)} | 0)}{\tilde{W}_N^{(j)}(\tilde{u}_{1:|\mathcal{M}|}^{(k)}, \hat{v}_{1:j-1}^{(k)} | 1)} \geq 1 \\ 1, & \text{otherwise.} \end{cases} \quad (3.8)$$

The output is the estimate of the information bits constructed in Phase 1 of encoding, reconstructing the erased bits in  $\widehat{\mathbf{U}}$ . Then the blocks corresponding to the degraded channel states, which are not decoded in the previous phase, can be decoded using the next phase.

### 3.1.3.3. Phase 3 – BSC Decoding 2

Finally, in Phase 3, the remaining blocks corresponding the degraded channel states can be decoded using the BSC( $p_1$ )-SC decoder with parameter  $p_1$ :

$$\hat{u}_i^{(b)} \triangleq \begin{cases} 0 & \text{if } i \in \mathcal{F} \\ \tilde{u}_i & \text{if } i \in \mathcal{M}, \\ d_{1,i}(y_{1:N}^{(b)}, \hat{u}_{1:i-1}^{(b)}) & \text{if } i \in \mathcal{G} \end{cases}, \quad i = 1:N \quad (3.9)$$

$$d_{1,i}(y_{1:N}^{(b)}, \hat{u}_{1:i-1}^{(b)}) \triangleq \begin{cases} 0, & \text{if } \frac{W_{1,N}^{(i)}(y_{1:N}^{(b)}, \hat{u}_{1:i-1}^{(b)} | 0)}{W_{1,N}^{(i)}(y_{1:N}^{(b)}, \hat{u}_{1:i-1}^{(b)} | 1)} \geq 1 \\ 1, & \text{otherwise.} \end{cases}$$

For the indices in set  $\mathcal{M}$ , the frozen bits are set to the values in  $\tilde{\mathbf{U}}$ , which is constructed from the estimate  $\hat{\mathbf{U}}$  from the previous phase, and for the indices in set  $\mathcal{F}$ , they are set to 0. For the indices in set  $\mathcal{G}$ , LLR calculations are done, just as if a classical BSC-SC was used. The overall decoding scheme is shown in Figure 3.3.

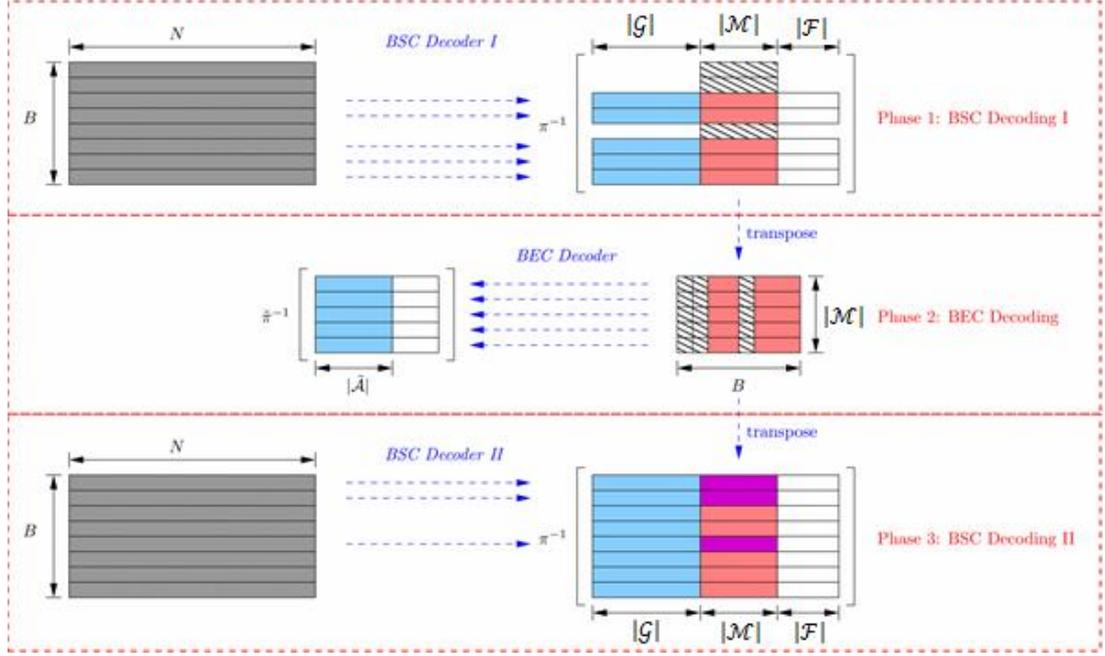
In Phase 1, only the blocks in the superior state are decoded; while in Phase 3, only the blocks in the degraded state are decoded. In Phase 2, all blocks are decoded using a BEC( $q_1$ )-SC decoder.

The performance of this scheme is intuitively discussed in [Si et al., 2014], finalized with a theorem stating that as long as the designed rates of polar codes do not exceed the corresponding channel capacities, all information bits are decoded with  $O(NB \log(NB))$  complexity.

### 3.2. Simulation Results

In this part, we compare the multi-phase polar coding scheme proposed by Si, Köylüoğlu and Vishwanath [Si et al., 2014] that we call the “SKV-code” with Arıkan’s original polar code [Arıkan, 2009], which we simply call “Arıkan’s polar code” or the “polar code”. We construct a polar code at the overall code rate  $q_1 R_1 + (1 - q_1) R_2$  with respect to both the degraded channel probability  $p_1$  and the superior channel probability  $p_2$  for performance comparison. The SKV-decoder is assumed to know the channel states, so that the appropriate blocks are decoded in Phase 1 and Phase 3 [Si et al., 2014]. Such information is redundant for Arıkan’s polar code, because the

selection of blocks “with respect to channel cases” does not exist, and the decoder works on all blocks irrespective of their fading states. In the simulations, the SKV-code and the Arikan’s polar code are subjected to exactly the same channel states and channel noise.



**Figure 3.3.** Polar decoder proposed in [Si et al., 2014] for a fading binary symmetric channel with two states (reproduced from [Si et al., 2014])

It should be noted that, from (3.4) and (3.5), one can see that the design parameters of the SKV-code result in code rates very close to the channel capacities which need to be decreased for practical code applications. Explicitly speaking, the BSC encoder’s code rate is  $R_1 = |G|/N = [1 - H(p_1) - \epsilon]$  whose channel capacity is  $C_{W_1} = 1 - H(p_1)$ ; and the code rate of the first BSC decoder (which is with respect to the good channel) is  $R_2 = [|\mathcal{G}| + |\mathcal{M}|]/N = [1 - H(p_2) - \epsilon]$  whose channel capacity is  $C_{W_2} = 1 - H(p_2)$ . Similarly, the code rate of the BEC encoder is  $R_3 = |\tilde{A}|/B = (1 - q_1 - \delta)$  and the channel capacity of  $\tilde{W} = BEC(q_1)$  is  $C_{\tilde{W}} = 1 - q_1$ . As can be

seen, proper offsets  $\epsilon$  and  $\delta$  are needed so that the code rates are smaller than the channel capacities.

We determine these offsets uniformly by fixing the code rate  $R$  to a percentage of the capacity  $C$ , such that  $R = \alpha C$ , where  $0 < \alpha < 1$ . This implies  $C_{W_i} - \epsilon = \alpha C_{W_i}$ ; hence,  $\epsilon = (1 - \alpha)C_{W_i}$  for the BSCs  $W_1$  and  $W_2$ ; similarly  $\delta = (1 - \alpha)C_{\tilde{W}}$  for the BEC,  $\tilde{W}$ .

- For the SKV-code;  $R_1 = K_1/N = \alpha C_{W_1}$ ,  $R_2 = K_2/N = \alpha C_{W_2}$  and  $R_{\tilde{W}} = K_3/B = \alpha C_{\tilde{W}}$ . As a result, the overall code rate,  $R = q_1 R_1 + (1 - q_1) R_2$  is settled as  $\alpha$  times the overall capacity  $C$  given by (3.3); i.e.,  $R = \alpha C$ .
  - For the BSCs  $W_1$  and  $W_2$ , the initial sets  $|\mathcal{G}|$ ,  $|\mathcal{M}|$  and  $|\mathcal{F}|$  given by (3.4) become  $|\mathcal{G}| = \alpha[1 - H(p_1)]N = K_1$ ,  $|\mathcal{M}| = \alpha[1 - H(p_2)]N - K_1 = K_2 - K_1$  and  $|\mathcal{F}| = N - K_2$ .
  - For the BEC  $\tilde{W}$ ,  $|\tilde{\mathcal{A}}| = \alpha(1 - q_1)B = K_3$ .
- For the polar code denoted by  $(N, K_p)$  the rate is also adjusted such that  $R_p = R = \alpha C$ .
  - Corresponding number of information bits of the polar code is found as  $K_p = q_1 K_1 + (1 - q_1) K_2 = R_p N$ .

In the figures of this section, we plot the BER performances of the SKV-codes and Arkan's polar codes versus  $q_1$ , the probability of being in the degraded channel; for a code length  $N = 256$  and number of blocks  $B = 256$ . On each figure, we add the uncoded BER performance,  $q_1 p_1 + (1 - q_1) p_2$ , as a green reference curve.

We fix the crossover probability of the degraded channel  $p_1$  to 0.1 and assign probabilities  $p_2 = 0.1, 0.05, 0.02, 0.01$  and  $0.001$  to the superior channel; to observe probability ratios of 1, 2, 5, 10, and 100, respectively. We start by choosing the coefficient  $\alpha = R/C$  as 0.6 for all values of  $p_1$  and  $p_2$ .

Firstly, the case of  $p_1 = p_2 = 0.1$  is simulated but not plotted, to confirm that the SKV and polar codes become the same when there is only one channel state. Specifically

speaking, the erasure channel  $\tilde{W}$  does not exist when  $p_1 = p_2$ , since the set  $|\mathcal{M}|$  shown in Figure 3.1 vanishes. As a result, there is no BEC encoder and decoder in the SKV-code. Its Phase 2 encoder is equal to the polar encoder; and its Phase 1 and Phase 3 decoders are equal to the polar decoder. Hence, the SKV and polar codes perform exactly the same when the crossover probabilities of the BSCs are equal, which is verified by the simulation results as well.

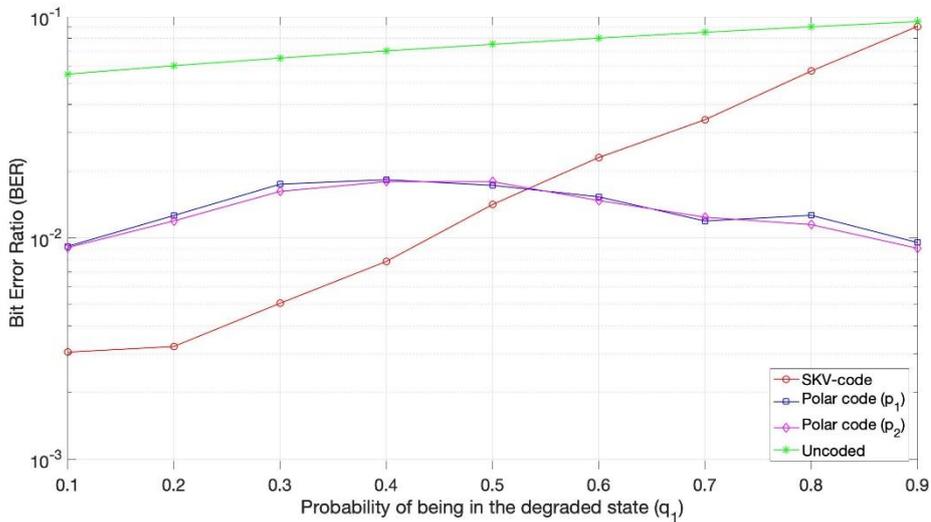
Comparison figures of BER performance begin with a superior channel of transition probability  $p_2 = 0.05$ ; i.e.,  $p_1/p_2 = 2$ . In Table 3.1, the channel capacities  $C_{W_1}$ ,  $C_{W_2}$ ,  $C_{\tilde{W}}$  and the overall channel capacity  $C$  corresponding to  $p_1 = 0.1$ ,  $p_2 = 0.05$  for different values of  $q_1$  are tabulated, along with the adjusted values for  $K_1$ ,  $K_2$ ,  $K_3$  and  $K_p$  which yield code rates equal to 60% of the related capacity.

**Table 3.1.** Channel capacities  $C_{W_1}$ ,  $C_{W_2}$ ,  $C_{\tilde{W}}$ , the overall channel capacity  $C$  for  $p_1 = 0.1$ ,  $p_2 = 0.05$ , corresponding number of information bits  $K_1$ ,  $K_2$ ,  $K_3$  and  $K_p$  for  $\alpha = 0.6$ .

$q_1$	$C_{W_1}$	$C_{W_2}$	$C(q_1)$	$C_{\tilde{W}}$	$K_1 =  \mathcal{G} $	$K_2 =  \mathcal{G}  +  \mathcal{M} $	$K_3 =  \tilde{\mathcal{A}} $	$K_p$
0.1	0.531	0.714	0.695	0.9	82	110	138	107
0.2			0.677	0.8			123	104
0.3			0.659	0.7			108	102
0.4			0.641	0.6			92	99
0.5			0.622	0.5			77	96
0.6			0.604	0.4			61	93
0.7			0.586	0.3			46	90
0.8			0.568	0.2			31	88
0.9			0.549	0.1			15	85

Notice that the maximum value of the capacity is  $C_{W_2} = 1 - H(p_2)$  for  $q_1 = 0$ ; i.e., when the channel is always in the superior state, and the minimum capacity value is  $C_{W_1} = 1 - H(p_1)$  for  $q_1 = 1$ , when the channel always remains in the degraded state. Therefore, the overall channel condition deteriorates as  $q_1$  increases.

In Figure 3.4, we plot the BER performance of the SKV-code with  $K_1 = 82$ ,  $|\mathcal{M}| = K_2 - K_1 = 28$  (red curve), and the polar codes designed both with respect to the degraded channel  $W_1$  (blue curve) and the superior channel  $W_2$  (pink curve), against increasing values of  $q_1$ . Recall that the overall code rate is a function of  $q_1$ , thus it does not remain constant at each step of  $q_1$ . This also yields varying polar code rates (from 0.43 to 0.32 as  $q_1$  goes from 0 to 1), as well. However, the code rates of the SKV-codes for the individual channels  $W_1$  and  $W_2$  are 0.32 and 0.43 respectively.



**Figure 3.4.** BER performance of the SKV-code and polar codes for different probabilities of the degraded state,  $N = B = 256$ ,  $p_1 = 0.1$ ,  $p_2 = 0.05$ , with code rates  $R_i = 0.6C_i$

From Figure 3.4, one observes that for  $q_1 < 0.53$ , the SKV-code performs much better than the polar codes. For the smallest value of  $q_1 = 0.1$ , the SKV-code provides 3 times better BER performance over the polar codes, and it is about 20 times better than the uncoded case. However, for  $q_1 > 0.53$ , the performance of the SKV-code starts to become worse than those of the polar codes and it approaches the uncoded BER performance at  $q_1 = 0.9$ ; where polar codes provide 10 times better BER performance over the uncoded case and the SKV-code. It is also observed that the

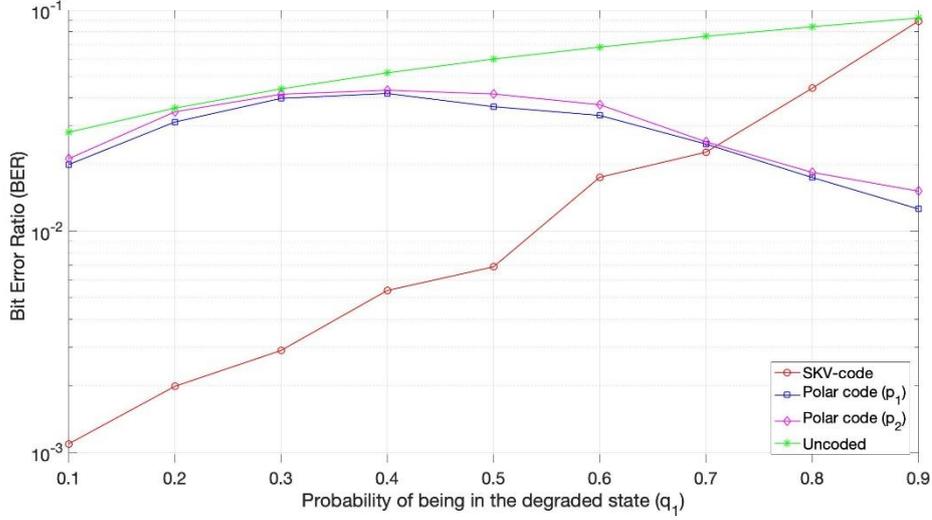
choice of the design parameter for the polar code (i.e.,  $p_1 = 0.1$  or  $p_2 = 0.05$ ) does not make an appreciable difference.

Next, we increase the gap between the channels and continue with  $p_1 = 0.1$  and  $p_2 = 0.02$ ; i.e.,  $p_1/p_2 = 5$ , that would yield the channel capacities  $C_{W_1}$ ,  $C_{W_2}$ ,  $C_{\bar{W}}$  and  $C$  along with the adjusted values for the information bits given in Table 3.2.

**Table 3.2.** Channel capacities  $C_{W_1}$ ,  $C_{W_2}$ ,  $C_{\bar{W}}$ , the overall channel capacity  $C$  for  $p_1 = 0.1$ ,  $p_2 = 0.02$ , corresponding number of information bits  $K_1$ ,  $K_2$ ,  $K_3$ ,  $K_p$  and rate  $R_p = 0.6C$ .

$q_1$	$C_{W_1}$	$C_{W_2}$	$C$	$C_{\bar{W}}$	$K_1$	$K_2$	$K_3$	$K_p$	$R_p$
0.1	0.531	0.859	0.826	0.9	82	132	138	127	0.496
0.2			0.793	0.8			123	122	0.477
0.3			0.760	0.7			108	117	0.457
0.4			0.728	0.6			92	112	0.438
0.5			0.695	0.5			77	107	0.418
0.6			0.662	0.4			61	102	0.398
0.7			0.629	0.3			46	97	0.379
0.8			0.597	0.2			31	92	0.359
0.9			0.564	0.1			15	87	0.340

Since the overall capacity is increased, the polar code rate  $R_p$  is also increased. We plot the BER performances in Figure 3.5, where one observes similar curve shapes to Figure 3.4: the SKV-code curve that is below the polar code curves for small  $q_1$  crosses them at  $q_1 = 0.71$ . Polar codes outperform the SKV code for  $q_1 > 0.71$ , but their performance is worse for small values of  $q_1$ , where the code rate  $R_p$  is higher. This declining behavior of polar codes for small  $q_1$  implies that, although the channel is more likely to be in the superior state, yet the errors made in a few degraded blocks by over-rate polar decoders dominate the overall number of errors. Again, the SKV-code approaches the uncoded BER curve as  $q_1$  increases.



**Figure 3.5.** BER performance of the SKV-code and polar codes for different probabilities of the degraded state,  $N = B = 256$ ,  $p_1 = 0.1$ ,  $p_2 = 0.02$ , with code rates  $R_i = 0.6C_i$

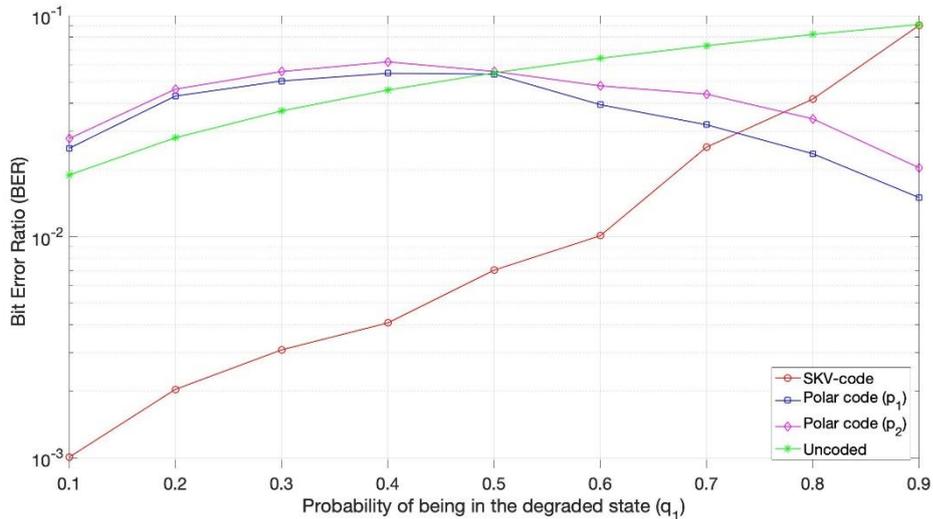
Decreasing  $p_2$  further to 0.01, we simulate a channel condition for  $p_1/p_2 = 10$ . In Table 3.3, we tabulate the corresponding channel capacities and the number of information bits for these channels.

**Table 3.3.** Channel capacities  $C_{W_1}, C_{W_2}, C_{\bar{W}}$ , the overall channel capacity  $C$  for  $p_1 = 0.1, p_2 = 0.01$ , corresponding number of information bits  $K_1, K_2, K_3, K_p$  and rate  $R_p = 0.6C$ .

$q_1$	$C_{W_1}$	$C_{W_2}$	$C$	$C_{\bar{W}}$	$K_1$	$K_2$	$K_3$	$K_p$	$R_p$
0.1	0.531	0.919	0.880	0.9	82	141	138	135	0.527
0.2			0.842	0.8			123	129	0.504
0.3			0.803	0.7			108	123	0.480
0.4			0.764	0.6			92	117	0.457
0.5			0.725	0.5			77	112	0.438
0.6			0.686	0.4			61	106	0.414
0.7			0.647	0.3			46	100	0.391
0.8			0.609	0.2			31	94	0.367
0.9			0.570	0.1			15	88	0.344

Once we plot the performance curves for these codes, as shown in Figure 3.6, we observe two distinct changes as compared to Figure 3.4 and Figure 3.5:

First, for smaller values of  $q_1$ , the BER performances of both polar codes are above the uncoded case, while the SKV-code remains more than 10 times better. The reason may be explained as follows: The channel is more likely to have a capacity of  $C_{W_2}$  for lower values of  $q_1$  and  $C_{W_1}$  for higher values of  $q_1$ . Since the polar codes do not use CSI (so they do not know which state the block is in), the code rate  $R_p$  should satisfy  $R_p < C_{W_1} < C_{W_2}$  so that it may decode correctly. Once we check this for  $q_1 = 0.1$  (see the last column of Table 3.3, shown by red), we notice that the code rate is  $135/256 = 0.527$ , which is very close to  $C_{W_1}$ . As a result, it is quite possible that degraded blocks cannot be decoded correctly. In addition, once such an error has been made, it is expected to be large; because with probability  $q_1$ , assuming half of the information bits are incorrectly decoded, polar BER has an additive component  $q_1/2$ , which is comparable to the uncoded BER of  $[q_1 p_1 + (1 - q_1) p_2]$ . Hence, one can reason the aforementioned dominance of degraded block errors over negligible amount of superior block errors.



**Figure 3.6.** BER performance of the SKV-code and polar codes for different probabilities of the degraded state,  $N = B = 256$ ,  $p_1 = 0.1$ ,  $p_2 = 0.01$ , with code rates  $R_i = 0.6C_i$

Second, for higher values of  $q_1$ , we now observe that the performances of two polar codes diverge from each other. This is reasonable because as  $q_1$  increases, the channel is more likely to be in the degraded state; and since the polar code designed for the degraded channel  $W_1$  performs better than the one designed for the superior channel  $W_2$ , the separation of BER curves is enhanced with increasing ratio of  $p_1/p_2$ . As a result, the pink curve starts to move away from the blue curve for  $q_1 > 0.5$ , and the blue curve appears as the best performance among all codes for  $q_1 > 0.73$ .

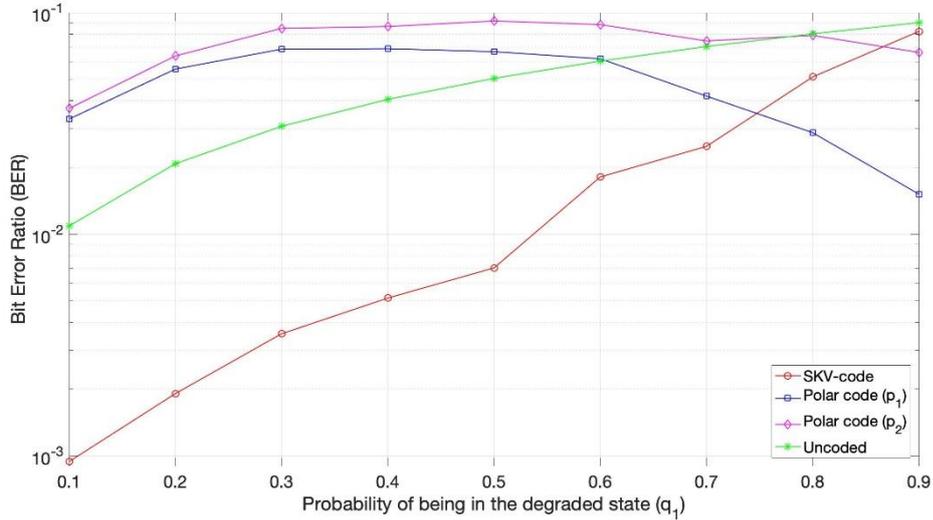
Apart from these observations, one may also notice the intersections of the SKV-code and polar code performances (which are similar to Figure 3.4 and Figure 3.5) at  $q_1 = 0.73$  for the blue curve of the polar code designed for the degraded channel, and at  $q_1 = 0.77$  for the pink curve of the polar code designed for the superior channel. At the largest value of  $q_1 = 0.9$ , the SKV-code again approaches the uncoded performance.

Next we pick a case where  $p_1/p_2 = 100$  by decreasing  $p_2$  to 0.001, while keeping  $p_1$  as 0.1. The parameters of this simulation are tabulated in Table 3.4. As the overall channel capacity is increased by using a better superior channel than before, the rates used in this simulation are the highest among the ones given in this chapter. Since the erasure channel  $\tilde{W}$  is a function of  $q_1$  only, there is no change in its rate  $R_3$ .

**Table 3.4.** Channel capacities  $C_{W_1}, C_{W_2}, C_{\tilde{W}}$ , the overall channel capacity  $C$  for  $p_1 = 0.1, p_2 = 0.001$ , corresponding number of information bits  $K_1, K_2, K_3, K_p$  and rate  $R_p = 0.6C$ .

$q_1$	$C_{W_1}$	$C_{W_2}$	$C$	$C_{\tilde{W}}$	$K_1$	$K_2$	$K_3$	$K_p$	$R_p$
0.1	0.531	0.989	0.943	0.9	82	152	138	145	0.566
0.2			0.897	0.8			123	138	0.539
0.3			0.851	0.7			108	131	0.512
0.4			0.806	0.6			92	124	0.484
0.5			0.760	0.5			77	117	0.457
0.6			0.714	0.4			61	110	0.430
0.7			0.668	0.3			46	103	0.402
0.8			0.623	0.2			31	96	0.375
0.9			0.577	0.1			15	89	0.348

In Figure 3.7, we plot the BER curves of the codes mentioned in Table 3.4. We witness the same two observations mentioned for Figure 3.6, but this time, the divergence of the two polar code curves is more pronounced as a result of the larger ratio of  $p_1/p_2$ . The polar code designed with respect to the good channel functions poorly, remaining almost always above the uncoded case except at  $q_1 = 0.9$ . When the channel is more likely to be in the degraded state, i.e., for  $q_1 \leq 0.5$ , we observe that the SKV-code performs approximately 10 times better than the uncoded case. Again, the SKV-code performs 30-35 times better than the polar codes for  $q_1 \leq 0.2$ .



**Figure 3.7.** BER performance of the SKV-code and polar codes for different probabilities of the degraded state,  $N = B = 256$ ,  $p_1 = 0.1$ ,  $p_2 = 0.001$ , with code rates  $R_i = 0.6C_i$

The red BER curve of the SKV-code intersects the blue curve of the polar code designed for the degraded channel at  $q_1 = 0.75$ , after which the polar code outperforms the SKV-code. For  $q_1 < 0.65$ , one observes that both of the polar codes perform worse than the uncoded case. Once the code rate of the polar code is checked from the last column of Table 3.4, one notices that the rate at small  $q_1$  values is not appropriate when the channel fades into the bad state. In other words, as  $q_1$  is small,

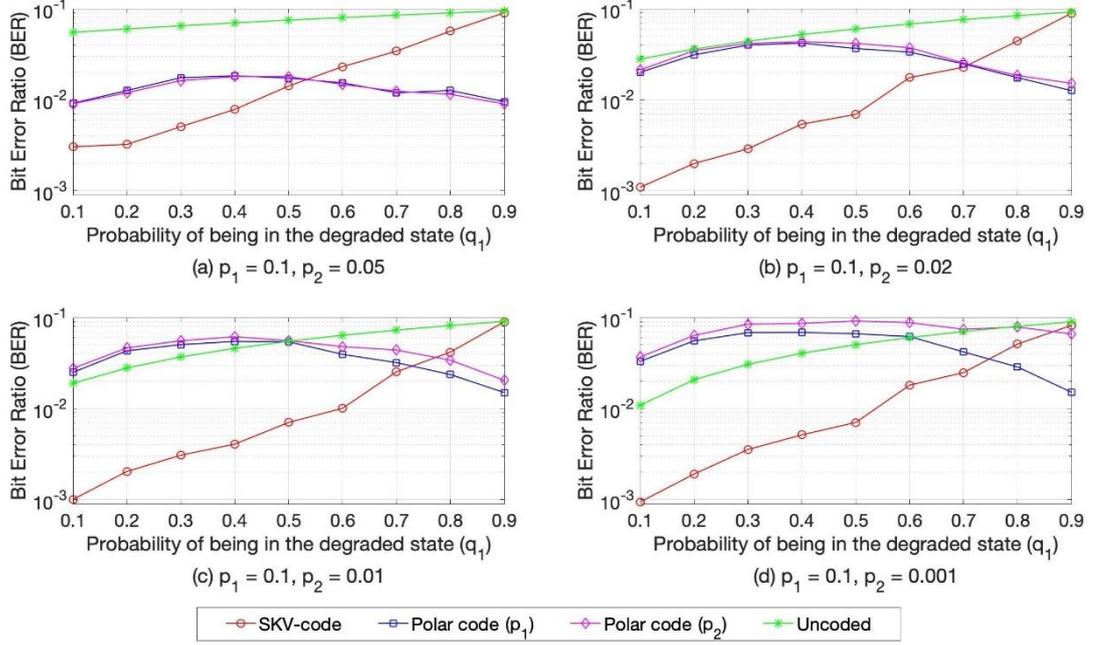
the code rate  $R_p$  is more likely to remain very close, or even above the channel capacity  $C_{W_1}$ . This is explicitly tabulated and shown in red in Table 3.5. Although this rate is suitable when the channel is in the good state; i.e.,  $R_p/C_{W_2}$  is suitable for all  $q_1$  values, the reason why the blue and pink curves remain above the uncoded curve is thus the rate being impractical for the degraded channel blocks.

**Table 3.5.** Ratio of the polar code rate  $R_p$  to capacities of the fading channels  $C_{W_1}$  and  $C_{W_2}$   
for  $p_1 = 0.1, p_2 = 0.001$

$q_1$	$C_{W_1}$	$C_{W_2}$	$C(q_1)$	$K_p$	$R_p/C_{W_1}$	$R_p/C_{W_2}$
0.1	0.531	0.989	0.943	145	1.066	0.572
0.2			0.897	138	1.015	0.545
0.3			0.851	131	0.964	0.518
0.4			0.806	124	0.911	0.489
0.5			0.760	117	0.861	0.462
0.6			0.714	110	0.810	0.435
0.7			0.668	103	0.757	0.406
0.8			0.623	96	0.706	0.379
0.9			0.577	89	0.655	0.352

In order to compare the above four cases where  $p_1/p_2$  takes the values 2, 5, 10 and 100, we combine the four figures, 3.4 to 3.7 in a single figure. Examining the BER performances of the SKV-code and polar codes in Figure 3.8, one observes that,

- SKV curve (of BER performance) remains almost the same in all four cases, with the rightmost end at  $q_1 = 1$  touching the uncoded BER of 0.1, but the polar code curves get worse as  $p_1/p_2$  and  $R_p = \alpha C$  increase.
- SKV-code is better than polar codes at small values of  $q_1$ , but it becomes worse for  $q_1 > p_0$ , and  $p_0$  increases from 0.53 to 0.75 with increasing  $p_1/p_2$ .
- Polar code designed for BSC( $p_1$ ) is the best solution for  $q_1 > 0.75$ , but  $R_p = \alpha C$  seems too high for polar codes, so  $\alpha$  needs to be properly decreased.



**Figure 3.8.** BER performances of the SKV-codes and polar codes for different probabilities of the degraded state,  $N = B = 256$ ,  $p_1 = 0.1$ ,  $p_2 =$  (a) 0.001, (b) 0.02, (c) 0.01 and (d) 0.001 with code rates  $R_i = 0.6C_i$

In the simulations discussed thus far, where  $q_1$  is changed for fixed values of  $p_1$  and  $p_2$ , variation of  $R$  with respect to  $q_1$  is decisive in the shape of the polar code BER performances. It is such that, as  $q_1$  increases from 0 to 1,  $R$  decreases from  $R_2$  to  $R_1$ . So, the choice of the coefficient  $\alpha = R/C$  seems to be crucial. While the constraint of  $R_p = R$ , where  $R = \alpha C = \alpha[q_1 C_{W_1} + (1 - q_1)C_{W_2}] = q_1 R_1 + (1 - q_1)R_2$ , forms a rate-equivalence between the SKV and single polar code simulations, one also needs to guarantee that  $R_p < C_{W_1}$ , so that the polar code can decode the degraded blocks. To satisfy  $R_p < C_{W_1} = R_1/\alpha$  is not easy at small values of  $q_1$ , where  $C$  approaches  $C_{W_2} > C_{W_1}$  and  $R_p$  approaches  $R_2 > R_1$ . Fulfillment of  $R_p < C_{W_1}$  puts a tighter restriction on  $\alpha = R_p/C$ , such that  $\alpha < (R_p/C_{W_1}) < 1$ . This inequality can be explained by employing the lower and upper bounds of  $C = q_1 C_{W_1} + (1 - q_1)C_{W_2}$  (that is  $C_{W_1} < C < C_{W_2}$  as  $1 > q_1 > 0$ ) on the ratio  $R_p/C$ . So, one obtains

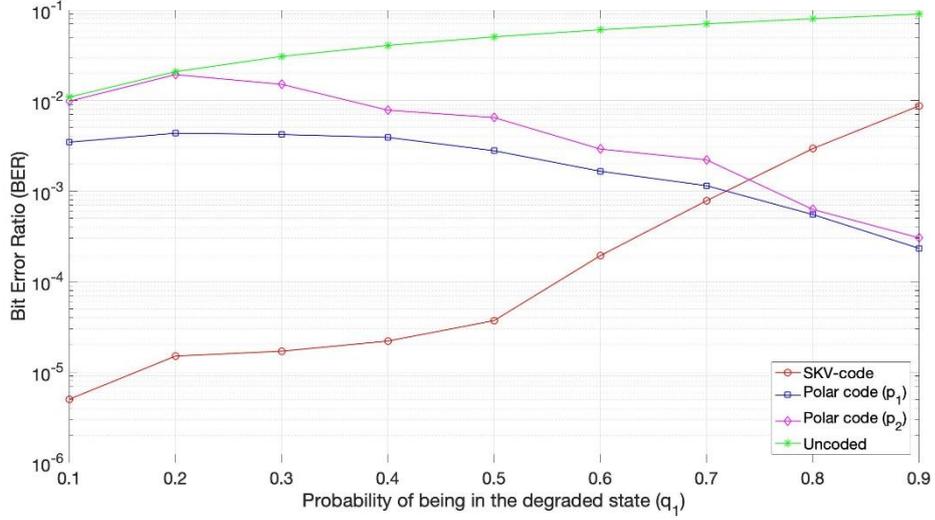
$$\frac{R_p}{C_{W_2}} < \alpha = \frac{R_p}{C} < \frac{R_p}{C_{W_1}} < 1$$

In the next simulation, we lower all code rates by choosing  $\alpha = 0.4$ . Among the previous  $p_1$ - $p_2$  pairs, we select the last one with  $p_2 = 0.001$ . New simulation parameters are given in Table 3.6, which shows that with  $\alpha = 0.4$ , the constraint of  $R_p < C_{W_1}$  is satisfied at small values of  $q_1$  as well; and all code rates remain in practical limits irrespective of the code used and the channel state.

**Table 3.6.** Channel capacities  $C_{W_1}, C_{W_2}, C_{\bar{W}}$ , the overall channel capacity  $C$  for  $p_1 = 0.1, p_2 = 0.001$ , corresponding number of information bits  $K_1, K_2, K_3, K_p$ , and rate  $R_p = 0.4C$ .

$q_1$	$C_{W_1}$	$C_{W_2}$	$C$	$C_{\bar{W}}$	$K_1$	$K_2$	$K_3$	$K_p$	$R_p$
0.1	0.531	0.989	0.943	0.9	54	101	92	96	0.375
0.2			0.897	0.8			82	92	0.359
0.3			0.851	0.7			72	87	0.340
0.4			0.806	0.6			61	82	0.320
0.5			0.760	0.5			51	78	0.305
0.6			0.714	0.4			41	73	0.285
0.7			0.668	0.3			31	68	0.266
0.8			0.623	0.2			20	63	0.246
0.9			0.577	0.1			10	59	0.230

BER performances of the SKV and polar codes for  $p_1 = 0.1$  and  $p_2 = 0.001$  at the overall code rate shown in the last column of Table 3.6 are plotted in Figure 3.9. The SKV-code achieves a BER as small as  $5 \times 10^{-6}$  at  $q_1 = 0.1$  and instead of touching the uncoded curve at  $q_1 = 0.9$ , it performs almost 10 times better. Polar code curves also remain below the uncoded case, because of the careful adjustment of all code rates.



**Figure 3.9.** BER performance of the SKV-code and polar codes for different probabilities of the degraded state,  $N = B = 256$ ,  $p_1 = 0.1$ ,  $p_2 = 0.001$ , with code rates  $R_i = 0.4C_i$

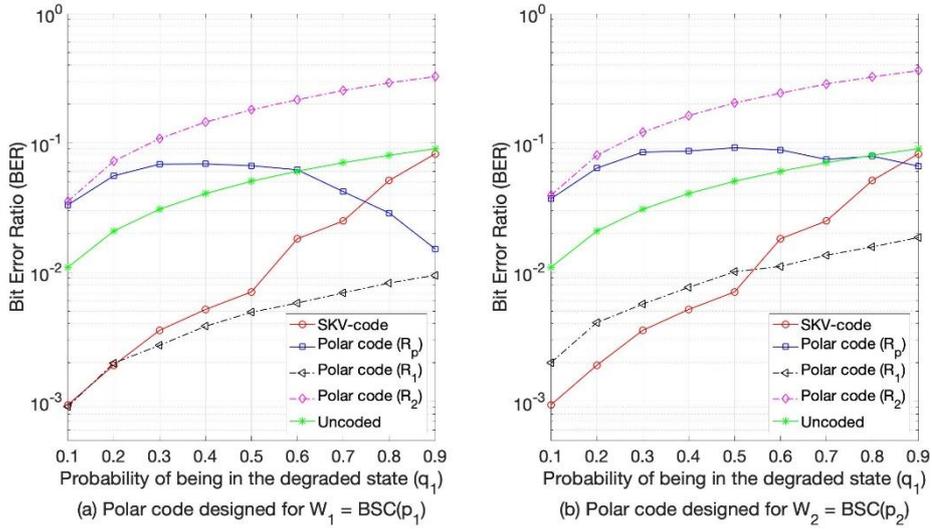
As  $q_1$ ; i.e., the percentage of the degraded blocks increases, the best choice is the polar code designed for the degraded channel (for  $q_1 > 0.72$  in this case). Polar code also has the advantage of not requiring any CSI, as opposed to the SKV-code. We find that the worsening of the SKV performance as  $q_1$  increases is mainly because of the errors made by the BEC decoder, as an inspection of Table 3.7 reveals.

**Table 3.7.** BER performance of the BEC decoder of the SKV-code for  $p_1 = 0.1$ ,  $p_2 = 0.001$ , and  $\alpha = R/C = 0.4$ .

$q_1$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
BER	0	0	0	0	0	0.000016	0.000087	0.000189	0.000361

One may also wonder how the original polar code would behave, if it were allowed to work at the rate bounds  $R_1$  and  $R_2$  rather than the fairly chosen code rate  $R_p$  that

increases from  $R_1$  to  $R_2$ , as  $q_1$  goes from 1 to 0. As expected and also spotted in Figure 3.10 (a), performances of the polar codes at rate  $R_1$  would produce lower bounds, and those at  $R_2$  would form upper bound to other BER performances.



**Figure 3.10.** BER performances of the SKV-code and polar codes for different probabilities of the degraded state,  $N = B = 256$ ,  $p_1 = 0.1$ ,  $p_2 = 0.001$ , with code rates  $R_i = 0.6C_i$ . Additionally, (a) polar codes are designed with respect to the degraded channel  $W_1$ , and (b) to the superior channel  $W_2$ , at rates  $R_p, R_1$  and  $R_2$ .

Finally, we express durations of some simulations using Intel Xeon CPU E5-1620 v3 @3.5 GHz, 32 GB RAM and 64-bit OS. For an  $N \times B$  block where  $N = B = 256$ , at  $q_1 = 0.1$ , a single encoding and decoding of the SKV-code takes about 6 seconds, while it lasts 3 seconds for the polar code on the average. At  $q_1 = 0.9$ , the SKV-code's duration increases to 21 seconds while that of the polar code remains the same. For the simulations in Figures 3.4 to 3.7, a simulation point is obtained in approximately 22 minutes at  $q_1 = 0.1$  and 52 minutes at  $q_1 = 0.9$ , where the decoding is performed over 120 channel realizations to measure BERs as low as  $10^{-3}$ , reliably. For the simulation in Figure 3.9, in order to obtain a reliable BER as low as  $5 \times 10^{-6}$ , the decoding is performed over 1000 channel realizations, which takes about roughly 48 hours.

## CHAPTER 4

### CHANNEL PARAMETERS AT WHICH POLAR CODES CONVERGE TO REED-MULLER CODES

$(N, K)$  polar codes introduced by Arıkan [Arıkan, 2008] are close cousins of  $(N, K)$  Reed-Muller (RM) codes, since they both have their generator matrix rows chosen from the rows of  $F^{\otimes n}$ , that is formed as the  $n^{\text{th}}$ -Kronecker product of the base matrix  $F = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ . Especially for small values of  $N$ , a meaningful portion of the  $K \times N$  generator matrices of the polar and RM codes are shared; i.e., they have many basis vectors in common. Arıkan and many other researchers have given insight regarding the relationship between RM and polar codes [Arıkan, 2008; Korada, 2009; Özgür, 2009], but we are mostly intrigued by the idea of constructing family of codes that interpolate between polar and RM codes of same rate and code length mentioned in [Mondelli et al., 2014] and [Mondelli, 2016]. Mondelli shows that polar codes designed for a specific channel are known to approach RM codes as the channel improves [Mondelli, 2016]. So if the channel is a BEC( $\epsilon$ ), the polar code converges to an RM code as channel's erasure rate  $\epsilon \rightarrow 0$ , and for the AWGN channel, the convergence occurs as the noise variance  $\sigma^2 \rightarrow 0$ , or equivalently, as the SNR increases. Akdođan has also noticed that for a BEC( $\epsilon$ ), as  $\epsilon$  becomes smaller, the generator matrix of the adaptive polar code becomes the same as that of the RM code [Akdođan, 2018].

In this chapter, we present some observations related to the convergence of polar codes to Reed-Muller codes. Motivated by improving the observation given for the  $(128, 64)$  codes in [Akdođan, 2018], we try to find out the values of  $\epsilon$  at which a polar code designed for a BEC( $\epsilon$ ) becomes an RM code, and present our observations for different code lengths and rates. We then extend this search to an AWGN channel, to

locate the values of the noise variance  $\sigma^2$ , at which the corresponding adaptive polar code converges to an RM code.

In Section 4.1, we review the convergence conditions of an adaptive polar code to an RM code, by replicating the related proposition and proof given by Mondelli [Mondelli, 2016], and attempting to clarify a few points. In the remaining part of the chapter, we construct  $(N, K)$  adaptive polar codes at several code lengths and rates for BECs with various  $\epsilon$ 's and AWGN channels with various SNRs. We then compare their generator matrices to those of the  $(N, K)$  RM codes with similar parameters. In Section 4.2, we detect the erasure rates  $\epsilon$  of BECs, where the generator matrix  $G_P(\epsilon)$  of the adaptive polar code designed for BEC( $\epsilon$ ) becomes the same as that of an RM code,  $G_{RM}$ . Similarly, in Section 4.3, we compute the SNR values of AWGN channels, where the generator matrix  $G_P(\text{SNR})$  of the adaptive polar code matches  $G_{RM}$  of the corresponding RM code.

#### 4.1. Convergence Proposition by Mondelli

In his Ph.D. dissertation, Marco Mondelli gives a proposition related to the convergence of a polar code designed for BEC( $\epsilon$ ) to a Reed-Muller code as  $\epsilon$  approaches to zero [Mondelli, 2016, Proposition 6.1]:

**“Proposition 6.1:** *An  $(N, K)$  Polar code designed for a BEC( $\epsilon$ ) channel becomes an RM code with the same rate and code length as  $\epsilon \rightarrow 0$ .”*

He then gives the following proof:

*“Suppose that the thesis is false, i.e., that we include  $g_{j^*}$ , but not  $g_{i^*}$ , with  $w_H(g_{i^*}) > w_H(g_{j^*})$ , where  $w_H(\cdot)$  denotes the Hamming weight. Since  $w_H(g_i) = 2^{\sum_{k=1}^n b_k^{(i)}} = 2^{w_H(b^{(i)})}$  for any  $i \in [N]$  from [Arikan, 2009, Proposition 17]), then  $w_H(b^{(i^*)}) > w_H(b^{(j^*)})$ .”*

From formula (6.1) (that is,  $Z_n^{(i)}(\epsilon) = f_{b_1^{(i)}} \circ f_{b_2^{(i)}} \circ \dots \circ f_{b_n^{(i)}}(\epsilon)$ ), we deduce that  $Z_n^{(i)}(\epsilon)$  is a polynomial in  $\epsilon$  with minimum degree equal to  $2^{w_H(\mathbf{b}^{(i)})}$ . Hence,

$$\lim_{\epsilon \rightarrow 0} \frac{Z_n^{(i^*)}(\epsilon)}{Z_n^{(j^*)}(\epsilon)} = 0,$$

which means that there exists  $\delta > 0$  such that for all  $\epsilon < \delta$ ,  $Z_n^{(i^*)}(\epsilon) < Z_n^{(j^*)}(\epsilon)$ .

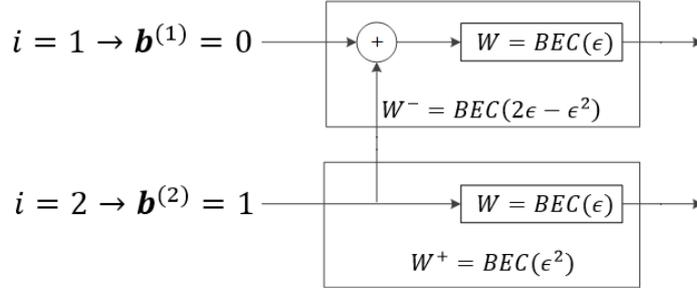
Consider a polar code designed for the transmission over a BEC( $\epsilon$ ), with  $\epsilon < \delta$ . Then, if this code includes  $g_{j^*}$ , it must also include  $g_{i^*}$ , which is a contradiction.”

Mondelli then proves that Proposition 6.1 not only holds for a BEC; it can be applied to any BMS channel [Mondelli, 2016]. Since the proof of Proposition 6.1 becomes clearer by understanding the mentioned equation,  $Z_n^{(i)}(x) = f_{b_1^{(i)}} \circ f_{b_2^{(i)}} \circ \dots \circ f_{b_n^{(i)}}(x)$ , we try to explain the parameters  $\mathbf{b}^{(i)} = b_n^{(i)}, f_{b_n^{(i)}}$  and  $Z_n^{(i)}(\epsilon)$  with a simple example which also points out some algebraic properties, such as bit representation of a polar code.

Input indices  $i$  of a length- $N$  polar code are integers in the range  $1 \leq i \leq N$ . The  $i^{\text{th}}$  input index is coupled with the  $n$ -bit vector  $\mathbf{b}^{(i)} = b_1^{(i)} b_2^{(i)} \dots b_n^{(i)}$  that is the binary expansion of  $(i - 1)$ , where  $n = \log_2 N$ .

Considering the simple case of  $N = 2$ , one may recall that after the first step of channel polarization, two versions of the channel  $W$ , namely  $W^-$  and  $W^+$  are obtained, with respective Bhattacharyya parameters  $2\epsilon - \epsilon^2$  and  $\epsilon^2$ , if  $W = \text{BEC}(\epsilon)$ .  $W^-$  is the bad channel denoted by “-” transform and  $W^+$  is the good channel denoted by “+” transform. For a code of length  $N = 2$ ,  $n = 1$ , so  $\mathbf{b}^{(i)}$  is a one-bit vector. In other

words, the input of  $W^-$  is mapped to  $\mathbf{b}^{(1)} = 0$  and the input of  $W^+$  is mapped to  $\mathbf{b}^{(2)} = 1$  as shown in Figure 4.1.



**Figure 4.1.** Polar code construction for  $N = 2$ .

From Figure 4.1, one observes the following:

- The binary representation of the 1<sup>st</sup> input is  $\mathbf{b}^{(1)} = 0$ . It can be related with a “-” transform and Bhattacharyya parameter  $f_{\mathbf{b}^{(1)}} = f_0 = 2\epsilon - \epsilon^2$ ,
- The binary representation of the 2<sup>nd</sup> input is  $\mathbf{b}^{(2)} = 1$ . It can be related with a “+” transform and Bhattacharyya parameter  $f_{\mathbf{b}^{(2)}} = f_1 = \epsilon^2$ .

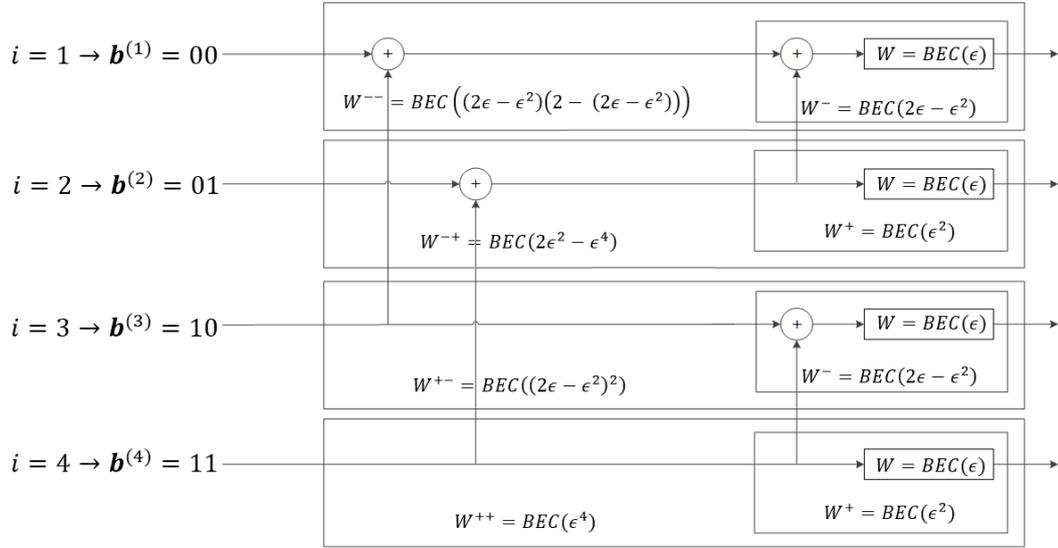
In other words, every 0 at the binary representation will be related with the “-” transform and Bhattacharyya parameter  $f_0$ , and similarly, every 1 at the binary representation will be related with the “+” transform and Bhattacharyya parameter  $f_1$ .

Now considering the polar code design for  $N = 4$ , Bhattacharyya parameters shown in Figure 4.2 are found as follows:

- For the worst channel  $W^{--}$ , which is coupled with  $i = 1$  and vector  $\mathbf{b}^{(1)} = 00$ ; one takes the initial Bhattacharyya parameter  $\epsilon$  and transforms it into  $f_0(\epsilon) = 2\epsilon - \epsilon^2$ , then substitutes this value again in  $2\epsilon - \epsilon^2$  to calculate  $f_0 \circ$

$f_0(\epsilon) = 2(2\epsilon - \epsilon^2) - (2\epsilon - \epsilon^2)^2$ , where “ $\circ$ ” denotes the function decomposition.

- Similarly, for the channel  $W^{-+}$ , which is coupled with  $i = 2$  and vector  $\mathbf{b}^{(2)} = 01$ ; one first computes  $f_1(\epsilon) = \epsilon^2$ , then substitutes this value in  $2\epsilon - \epsilon^2$  to calculate  $f_0 \circ f_1(\epsilon) = 2\epsilon^2 - \epsilon^4$ .
- For the channel  $W^{+-}$ ,  $i = 3$  and  $\mathbf{b}^{(3)} = 10$ ; so one first computes  $f_0(\epsilon) = 2\epsilon - \epsilon^2$ , then substitutes it in  $\epsilon^2$  to calculate  $f_1 \circ f_0(\epsilon) = (2\epsilon - \epsilon^2)^2$ .
- For the channel  $W^{++}$ ,  $i = 4$  and  $\mathbf{b}^{(4)} = 11$ ; so one first computes  $f_1(\epsilon) = \epsilon^2$ , then substitutes this value in  $\epsilon^2$  again, to calculate  $f_1 \circ f_1(\epsilon) = \epsilon^4$ .



**Figure 4.2.** Polar code construction for  $N = 4$ .

Due to the iterative construction of polar codes, the same procedure is repeated by adding one more bit to  $\mathbf{b}^{(i)}$ , as  $N$  doubles. As a result, the Bhattacharyya parameter of the  $i^{\text{th}}$  input is calculated as

$$Z_n^{(i)}(x) = f_{b_1^{(i)}} \circ f_{b_2^{(i)}} \circ \dots \circ f_{b_n^{(i)}}(x) \quad (4.1)$$

where  $f_0(x) \triangleq 2x - x^2$ ,  $f_1(x) \triangleq x^2$  and the  $n$ -bit vector  $\mathbf{b}^{(i)} = b_1^{(i)} b_2^{(i)} \dots b_n^{(i)}$ .

Proposition (6.1) also uses the fact that  $Z_n^{(i)}(x)$  is a polynomial with minimum degree of  $2^{w_H(\mathbf{b}^{(i)})}$  (See the proof in Appendix A.2). This formula can be applied to any binary input-DMS channel, where  $x$  is the initial Bhattacharyya parameter of the channel, as described in Chapter 2. In order to construct a  $(2^n, K)$  polar code, one simply selects  $K$  rows of the  $n^{\text{th}}$  Kronecker product matrix  $F^{\otimes n}$ , which minimize (4.1).

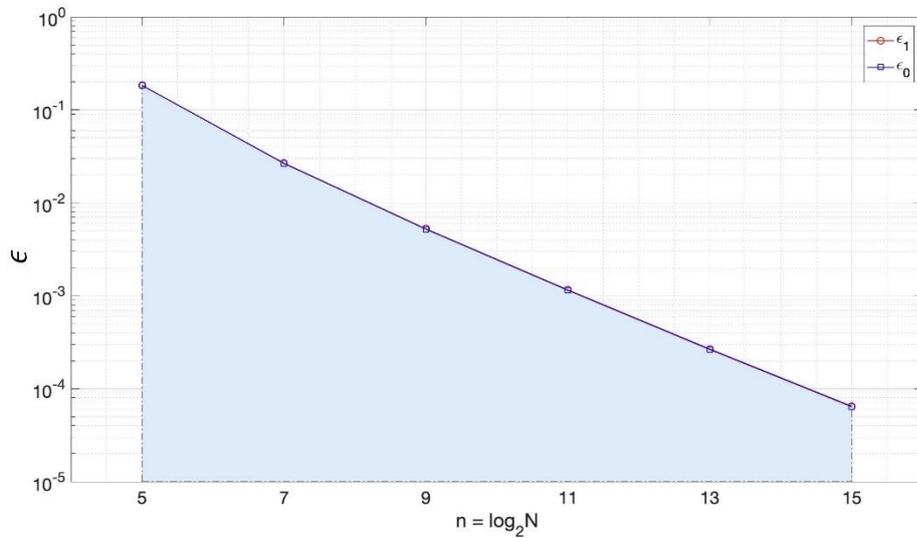
The proof of Proposition 6.1 given by Mondelli (and repeated at the beginning of this section), affirms the convergence of polar codes to RM codes as  $\epsilon$  approaches 0. In the light of this result, we calculate some numerical values of design erasure rates or design SNRs respectively, required by the convergence of polar codes to RM codes, for the BEC in Section 4.2, and for the AWGN channel in Section 4.3.

## 4.2. BEC Erasure Rates for the Convergence of Polar Codes to RM Codes

We firstly design  $(N, N/2) = (2^n, 2^{n-1})$  adaptive polar codes for the BEC( $\epsilon$ ) at various erasure rates  $\epsilon$  that differ in small steps of  $\Delta\epsilon$ . Since the RM codes of rate 0.5 only exist for odd values of  $n = \log_2 N$ , we use the odd values,  $n = 5, 7, 9, 11, 13, 15$  (and codeword lengths of  $N = 32, 128, 512, 2048, 8192, 32768$ ) in our generator matrix computations. We compare the generator matrix of the resulting polar code to that of the  $(N, N/2)$  RM code in order to detect  $\epsilon_0$ , that we define as the erasure rate, below which the two generator matrices become exactly the same. Similarly, we define  $\epsilon_1$  as the erasure rate, above which two generator matrices start to differ by one basis vector. Calling the step size  $\Delta\epsilon$ ,  $\epsilon_1 = \epsilon_0 + \Delta\epsilon$ .

In Figure 4.3, we plot the values of  $\epsilon_0$  at which generator matrices of polar and RM codes of rate 0.5 are the same (together with  $\epsilon_1$ ), with respect to different values of  $n = \log_2 N$ . As the code length  $N$  increases, “the value of  $\epsilon$  at which the polar and

RM codes converge to each other” decreases; therefore, the chosen value of  $\Delta\epsilon$  that separates  $\epsilon_0$  and  $\epsilon_1$  needs to be suitably reduced. At each increase of  $n$ , we reduce the step size by 4, since the number of generator matrix rows  $K = 2^{n-1}$  quadruples. Notice that curves of  $\epsilon_0$  and  $\epsilon_1$  that differ by  $\Delta\epsilon$  seem very close to each other. The blue shaded region that remains below the  $\epsilon_0$ -curve is the region of erasure rates, where the  $(2^n, 2^{n-1})$  polar codes converge to RM codes.



**Figure 4.3.** Variation of design erasure probabilities  $\epsilon_0$  and  $\epsilon_1$  versus  $n$ , for which  $(2^n, 2^{n-1})$  adaptive polar codes designed for BEC( $\epsilon$ ) converge to RM codes.

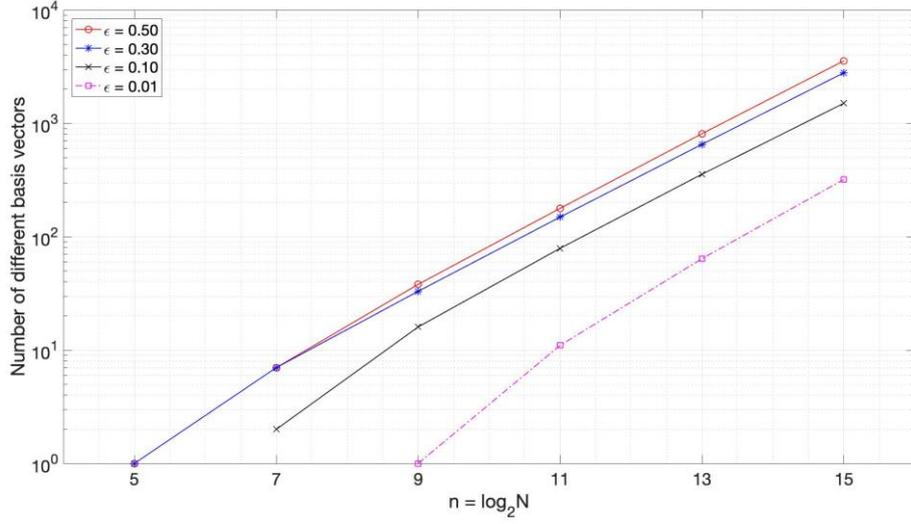
In Table 4.1, we highlight some details of Figure 4.3, where one may observe that the generator matrix of the  $(32, 16)$  adaptive polar designed for a BEC with erasure rate  $\epsilon_0 = 0.183$ , is the same as the generator matrix of the  $(32, 16)$  RM code; whereas for the polar code designed for  $\epsilon_1 = 0.184$ , the generator matrices differ by one basis vector, the remaining 15 basis vectors remaining the same. Precision of the computed  $\epsilon_0$  and  $\epsilon_1$  values depends on the chosen step size  $\Delta\epsilon$  shown in Table 4.1. For example, the polar code of length  $N = 512$  becomes an RM code roughly for  $\epsilon \leq 0.005$ , while for  $N = 2048$  this happens for  $\epsilon \leq 0.0011$ . For  $N = 8192$ , we need to decrease the

step size to  $4 \times 10^{-6}$  in order to observe the convergence for  $\epsilon \leq 0.00026$ . Clearly, if one uses smaller step sizes,  $\epsilon_0$  and  $\epsilon_1$  values can be obtained with larger precision.

**Table 4.1.** Erasure probabilities  $\epsilon_0$  and  $\epsilon_1$ , for which  $(N, N/2) = (2^n, 2^{n-1})$  adaptive polar codes designed for BEC( $\epsilon$ ) converge to RM codes.

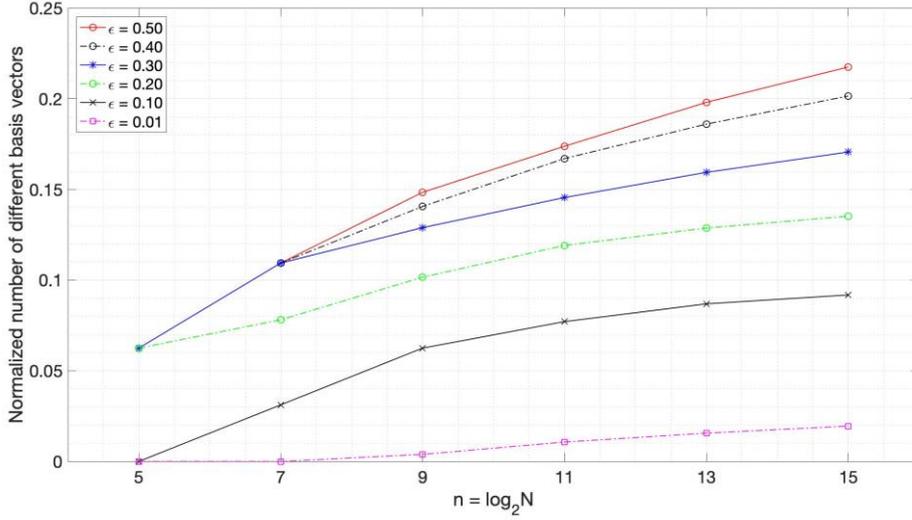
$N$	32	128	512	2048	8192	32768
$n = \log_2 N$	5	7	9	11	13	15
$\epsilon_0$	0.183	0.02650	0.00520	0.001145	0.000264	0.000064
$\epsilon_1$	0.184	0.02675	0.00526	0.001160	0.000268	0.000065
$\Delta\epsilon$	0.001	0.00025	0.00006	0.000015	0.000004	0.000001

Recall that in Figure 4.3, the region above the given curves corresponds to where the generator matrices of polar and RM codes differ. They start to differ by one basis vector at  $\epsilon = \epsilon_1$  but as the design parameter  $\epsilon$  of the polar code increases, the number of different basis vectors also increases. Defining  $D(n, R, \epsilon)$  as the “number of different rows between the generator matrices of rate- $R$ ,  $(2^n, K)$  RM codes and adaptive polar codes designed for BEC( $\epsilon$ )”; in Figure 4.4, we plot  $D(n, 0.5, \epsilon)$  versus  $n$ , on a logarithmic scale since  $D(n, 0.5, \epsilon)$  grows exponentially with  $n$ , for  $\epsilon = 0.01, 0.1, 0.3$  and  $0.5$ . One observes that as  $\epsilon$  gets smaller,  $D(n, 0.5, \epsilon)$  for a fixed  $n$  also gets smaller as predicted [Mondelli, 2016, Proposition 6.1].



**Figure 4.4.** Number of different generator matrix rows  $D(n, 0.5, \epsilon)$  of the  $(2^n, 2^{n-1})$  RM and adaptive polar codes designed for BEC( $\epsilon$ ) versus  $n$ , for various design erasure rates.

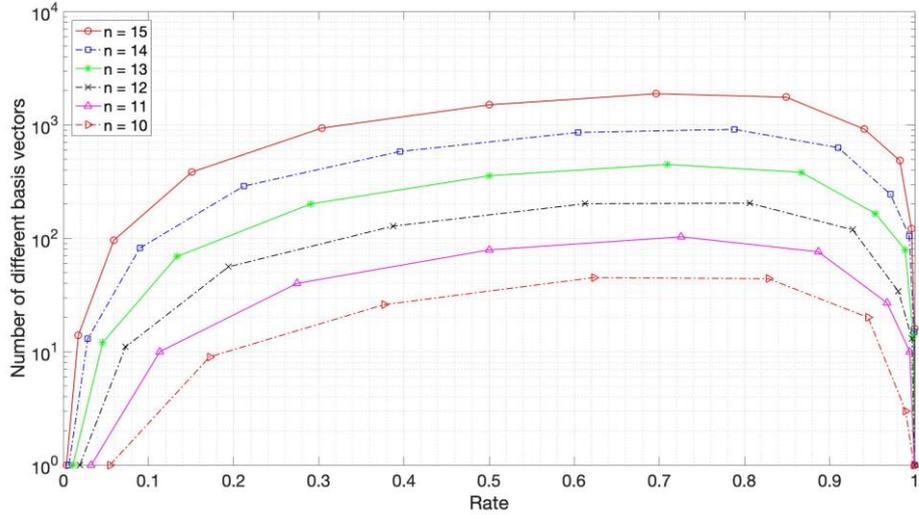
The exponential increase of  $D(n, R, \epsilon)$  with  $n$  is expected, since the total number of generator matrix rows is also increasing exponentially with  $n$  ( $K = 2^{n-1}$ ). To observe the difference percentage between the basis vectors of polar and RM codes, we normalize  $D(n, 0.5, \epsilon)$  by  $K$ , and plot  $D(n, 0.5, \epsilon)/K$  versus  $n$  in Figure 4.5, for polar codes designed over BEC( $\epsilon$ )'s with various erasure rates  $\epsilon = 0.01, 0.1, 0.2, 0.3, 0.4$  and  $0.5$ . The vertical axis of Figure 4.5 now doesn't need to be logarithmically scaled as that of Figure 4.4, but it is linear. The curves of Figure 4.5 show that for  $\epsilon \leq 0.5$  and  $n \leq 15$ , the normalized ratio of differing rows between the generator matrices of rate-0.5 RM codes and polar codes designed for BEC( $\epsilon$ ) remains less than 22%; i.e.,  $D(n, 0.5, \epsilon)/K \leq 0.22$ . Examining the shapes of these curves, one may also conjecture that as  $n$  grows further for a fixed value of the design erasure rate,  $\epsilon$ , the normalized number of different rows between two generator matrices approaches an asymptote.



**Figure 4.5.** Normalized ratio of different rows,  $D(n, 0.5, \epsilon)/K$ , between the generator matrices of the  $(2^n, 2^{n-1})$  RM and adaptive polar codes designed for BEC( $\epsilon$ ) versus  $n$ , for various erasure rates.

The capacity of BEC( $\epsilon$ ) is  $C = 1 - \epsilon$ , and one needs to choose  $R < C$  for reliable transmission; hence  $R = 0.5$  codes are not suitable for a BEC( $\epsilon$ ) when  $\epsilon \geq 0.5$ , and this is the reason for their exclusion from Figure 4.5. On the other hand, the smallest value of  $\epsilon$  used in Figure 4.5 corresponds to a channel capacity as large as 0.99. So, we also study codes at different rates,  $R$ , within the rate restrictions of the RM code design, where one defines the order  $r$  and selects the rate accordingly. As described in Chapter 2, the order  $r$  of an RM code of length  $N = 2^n$  can be selected as  $1 \leq r \leq n - 1$ , for any  $n$ . The information word length is then  $K = \sum_{i=0}^r \binom{n}{i}$  with the corresponding rate,  $R = K/N = \sum_{i=0}^r \binom{n}{i} / 2^n$ . For example, if  $r = 5$  and the code length is chosen as  $N = 2^{14} = 16384$ , corresponding RM code shown in  $\left\{ \begin{smallmatrix} n \\ r \end{smallmatrix} \right\}$  notation as  $\left\{ \begin{smallmatrix} 14 \\ 5 \end{smallmatrix} \right\}$ , and in  $(N, K)$  notation as  $(16384, 3473)$  has the rate 0.21. Since the polar codes can be constructed at any rate; we restrict their rates to those of the  $(N, K)$  RM codes.

In the remaining part of this section, we explore the dependence of  $D(n, R, \epsilon)$  on the code rate  $R$  for various channels with different erasure probabilities  $\epsilon$ . We start by fixing  $\epsilon$  to 0.1 and observing the number of different rows,  $D(n, R, 0.1)$ , between the generator matrices of RM and polar codes designed for BEC(0.1), at various rates. Different from the experiments performed previously at the fixed rate of 0.5, RM codes can now be constructed for even values of  $n$  as well. In Figure 4.6, we plot  $D(n, R, 0.1)$  versus rate  $R$ , for  $10 \leq n \leq 15$ . Note that code rates above 0.9 are not practically meaningful since they exceed the channel capacity of BEC(0.1).



**Figure 4.6.** Number of differing rows  $D(n, R, 0.1)$  between the generator matrices of the adaptive polar codes designed for BEC(0.1) and RM codes at various rates.

One observes from Figure 4.6 that:

- i. For  $n_1 < n_2$ ,  $D(n_1, R, 0.1) < D(n_2, R, 0.1)$ . This is quite reasonable because the number of basis vectors  $K = 2^n R$  also increases as  $n$  increases.
- ii. All curves are concave functions of the rate  $R$ . Denoting  $R_{0.1,n}$  as the rate where  $D(n, R, 0.1)$  is maximized for each  $n$ , each curve increases if  $0 < R <$

$R_{0.1,n}$  and decreases if  $R_{0.1,n} < R < 1$ . Concavity of the curves can be understood by the following argument: Basis vectors of the  $(N, K)$ , rate- $R$  polar and RM codes are both chosen from the same set; i.e.,  $N$  rows of  $F^{\otimes n}$  formed as the  $n^{\text{th}}$ -Kronecker product of the base matrix  $F = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ . If  $D$  simply denotes the number of different rows between the generator matrices  $G_{RM}$  and  $G_P$  of these two codes (that is  $D(n, R, \epsilon)$  defined on page 62), and  $S$  denotes the number of the same rows of  $G_{RM}$  and  $G_P$ , then  $D + S = K$ . Since rows of both  $G_{RM}$  and  $G_P$  are chosen from  $F^{\otimes n}$ , when they differ by  $D$  rows,  $2D$  rows of  $F^{\otimes n}$  out of  $N$  have been used. Also considering the additional  $S$  similar rows; one concludes that  $2D + S \leq N$ . Combining these two restrictions, we obtain,

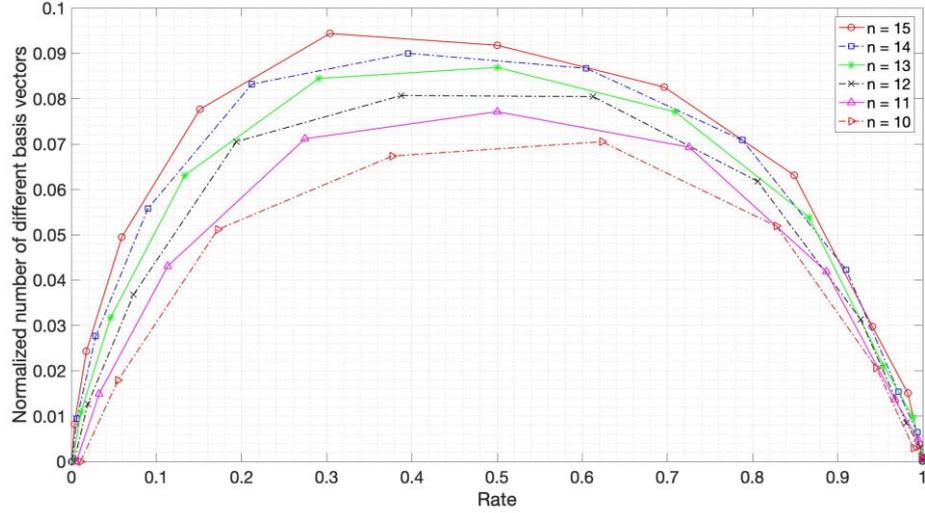
$$\begin{aligned} 2D + S &\leq N \\ D + S &= K \end{aligned} \implies K - S = D \leq N - K \quad (4.2)$$

Dividing (4.2) by  $N$ , it is seen that  $D/N$  satisfies

$$R - (S/N) = D/N \leq 1 - R \quad (4.3)$$

So, for a fixed codeword length  $N$ , the initial growth of  $D$  with  $R$  for  $0 < R < R_{0.1,n}$ , is explained by the equality on the left side of (4.3), and its decline for  $R_{0.1,n} < R < 1$  is justified by the inequality on the right side of (4.3).

We then normalize the number of different rows shown in Figure 4.6 by dividing it with the respective number of information bits  $K$ , and plot  $D(n, R, 0.1)/K$  in Figure 4.7. Again we note that code rates above 0.9 are not practically meaningful since they exceed the channel capacity of BEC(0.1).



**Figure 4.7.** Normalized number  $D(n, R, 0.1)/K$  of differing rows between the generator matrices of the adaptive polar codes designed for BEC(0.1) and RM codes, versus the code rate  $R$ .

Previously in Figure 4.6, rate values  $R_{0.1,n}$  that maximize the curves of have been observed in the range (0.7, 0.85). However, once these numbers are normalized by  $K$  as in Figure 4.7, peaks of the curves shift to the left and maxima start to occur in the rate range (0.3, 0.62); because as  $R$  increases, the normalization parameter  $K$  also increases. One observes almost a symmetrical shape in the normalized values,  $D(n, R, 0.1)/K$ .

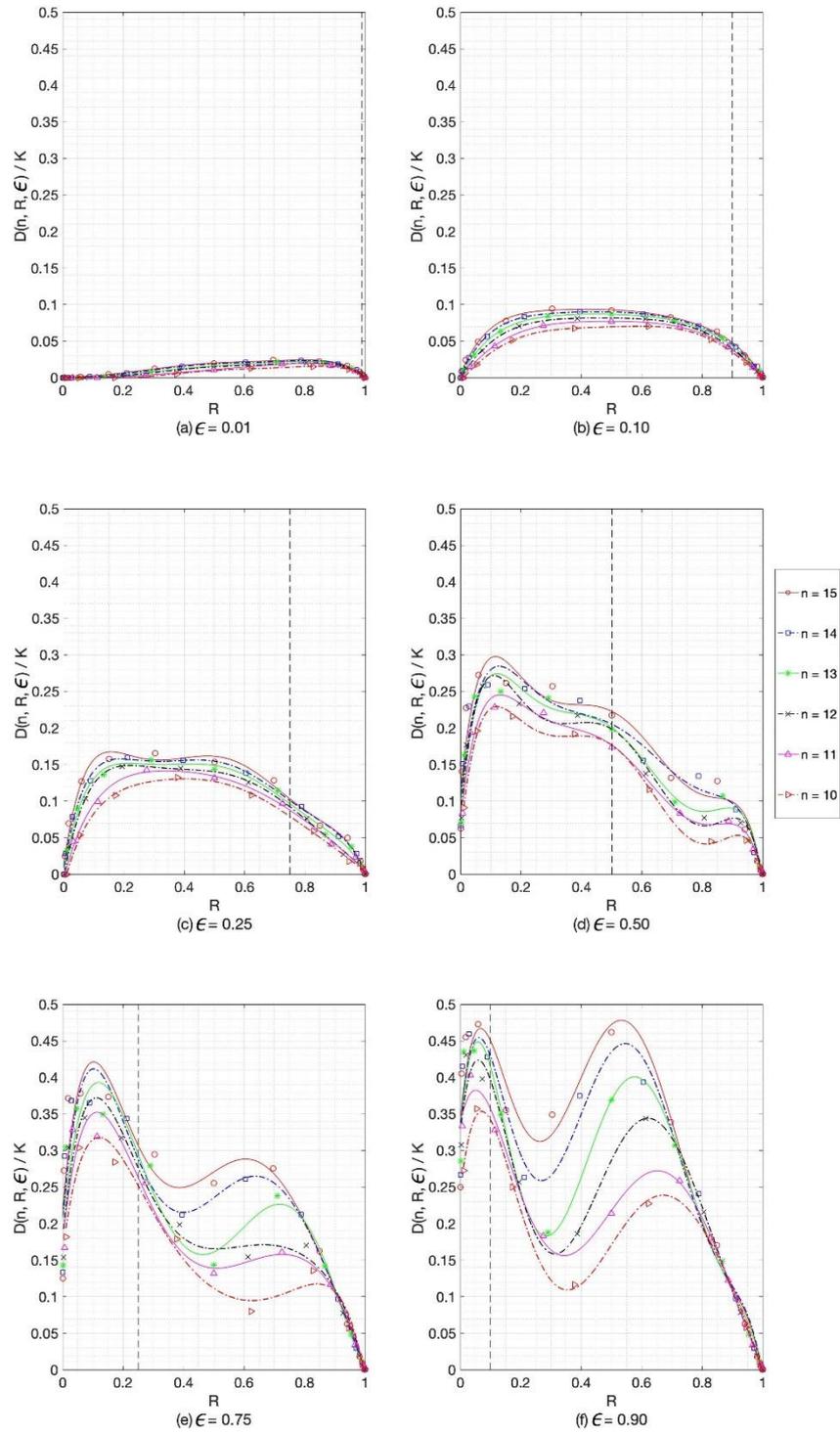
Recalling the definition of  $D(n, R, \epsilon)$  as the “number of different rows between the generator matrices of rate- $R$ ,  $(2^n, K)$  RM codes and adaptive polar codes designed for BEC( $\epsilon$ )”, Figure 4.7 also shows that normalized  $D(n, R, 0.1)/K$  values increase with  $n$ , but remain below 9.5% for all  $n \leq 15$ . They are also observed to decrease with  $R$  for  $R > 0.5$ ; so the previous curve  $D(n, 0.5, 0.1)/K$  given in Figure 4.5 seems to form an upper bound for  $D(n, R, 0.1)/K$  curves of Figure 4.7, when  $R > 0.5$ .

So, as a final concern of this section, we compute the normalized numbers  $D(n, R, \epsilon)/K$  (of different generator rows of the RM codes and adaptive polar codes designed for BEC( $\epsilon$ )’s) versus  $R$ , considering different erasure rates  $\epsilon = 0.01, 0.1$ ,

0.25, 0.5, 0.75 and 0.90. We plot the results in Figure 4.8, also indicating the related channel capacities  $(1 - \epsilon)$  on each subfigure with a vertical dashed line, where the proper rates for reliable communication remain on the left. For practical reasons, this rate should not be too far away from capacity, as well.

Comparing the six design- $\epsilon$  values of Figure 4.8, one may observe that for  $n \leq 15$ , the normalized ratio of differing rows between the generator matrices of RM codes and polar codes designed for BEC( $\epsilon$ ),  $D(n, R, \epsilon)/K$  satisfies:

- i.* For  $n, R$  fixed, and  $\epsilon_1 < \epsilon_2$ ,  $[D(n, R, \epsilon_1)/K] < [D(n, R, \epsilon_2)/K]$ : This is an expected result [Mondelli, 2016, Proposition 6.1] since polar codes converge to RM codes as  $\epsilon \rightarrow 0$ . For all  $n \leq 15$ , the maximum of  $D(n, R, \epsilon)/K$  is found to be 47% at  $\epsilon = 0.9$ , which decreases to 43% at  $\epsilon = 0.75$ , 28% at  $\epsilon = 0.5$ , 17% at  $\epsilon = 0.25$ , 9.5% at  $\epsilon = 0.1$ , and 2.5% at  $\epsilon = 0.01$ . At meaningful rates (i.e.,  $R < C$ ), the maximum of  $D(n, R, \epsilon)/K$  is found to be 47% at  $\epsilon = 0.9$ .
- ii.* For small values of  $\epsilon$ ,  $D(n, R, \epsilon)/K$  curves seem to be piecewise concave functions of the rate  $R$ , where there is a point  $\bar{R}_{\epsilon, n} < C$  for each  $n$  and  $\epsilon$  such that  $D(n, R, \epsilon)/K$  is maximized. For higher values of  $\epsilon$ , such concave-like behavior is only observed at meaningful transmission rates less than capacity; and for  $R > C$ , a second peak starts to develop, as becomes especially apparent at  $\epsilon = 0.9$ . We still continue to denote  $\bar{R}_{\epsilon, n} < C$ ; i.e., the first peak, as the rate value for which  $D(n, R, \epsilon)/K$  is maximum for meaningful rates  $R < C$ .
- iii.* As the design erasure rate  $\epsilon$  increases, we observe that  $\bar{R}_{\epsilon, n}$  decreases; i.e., the peaks of  $D(n, R, \epsilon)/K$  versus  $R$  curves shift from right to left and start to occur at lower rates. Specifically, for  $\epsilon = 0.01$ , one can see that the maximum value of  $D(n, R, \epsilon)/K$  occurs for code rates in the (0.7, 0.9) range; while for  $\epsilon = 0.1$ ,  $\bar{R}_{\epsilon, n}$  is within the interval (0.3, 0.62); for  $\epsilon = 0.25$ , it is in the range (0.2, 0.3);



**Figure 4.8.** Normalized number  $D(n, R, \epsilon)/K$  of differing rows between the generator matrices of the adaptive polar codes designed for BEC( $\epsilon$ ) and RM codes, versus the code rate  $R$ .

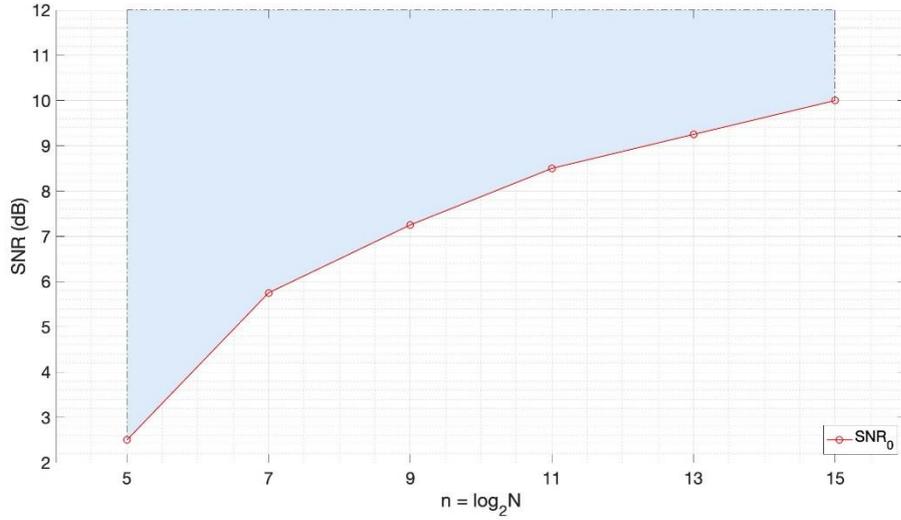
for  $\epsilon = 0.5$  and  $0.75$ , it is around  $R = 0.1$  and  $\bar{R}_{\epsilon,n}$  becomes less than  $0.1$  for  $\epsilon = 0.9$ .

In short, for  $\epsilon_1 < \epsilon_2$ , one concludes that  $\bar{R}_{\epsilon_1,n} > \bar{R}_{\epsilon_2,n}$ , where  $\bar{R}_{\epsilon,n}$  is the rate value at which  $D(n, R, \epsilon)/K$  is maximum for meaningful rates  $R < C$ .

### 4.3. AWGN Channel SNRs for the Convergence of Polar Codes to RM Codes

In this section, it is aimed to observe the practical conditions for the convergence of polar codes to RM codes for the binary input-additive white Gaussian (BI-AWGN) channels. Knowing that the convergence occurs as the channel approaches to a perfect one, and since the Bhattacharyya parameter of AWGN channel is  $Z_{\text{AWGN}}(W) = e^{-E_s/\sigma^2}$  as mentioned in Table 2.1 of Chapter 2, we search for the SNR values for which the convergence would occur. We compute our results in terms of the SNR =  $10 \log_{10}(E_s/\sigma^2)$  values, where  $E_s$  is the signal power.

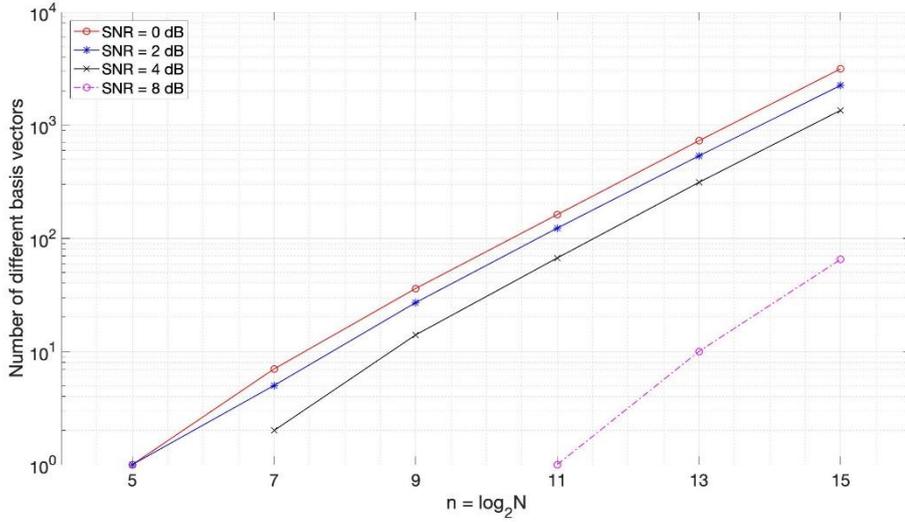
We firstly construct  $(N, N/2) = (2^n, 2^{n-1})$  adaptive polar codes over the AWGN channel with respect to different channel SNRs and compare each generator matrix  $G_P(\text{SNR})$  with the generator matrix  $G_{RM}$  of the corresponding  $(2^n, 2^{n-1})$  RM code. We use the same code lengths as before and increase the SNR from 0 dB to 15 dB, with a step size of 0.25 dB. Similar to the BEC case, we look for the SNR values at which polar and RM codes of rate 0.5 start to differ. Calling the SNR value that they become the same as  $\text{SNR}_0$ , we plot these values versus  $n$  in Figure 4.9. As in Figure 4.3 of Section 4.2, we indicate the region where the  $(2^n, 2^{n-1})$  polar codes converge to RM codes by the blue-shaded area, which now remains above the  $\text{SNR}_0$  curve. We note that polar codes designed for  $\text{SNR}_1 = \text{SNR}_0 - 0.25$ , differ by one basis vector from those of the RM codes at all values of  $n$ .



**Figure 4.9.** Variation of the design SNR values versus  $n$ , for which  $(2^n, 2^{n-1})$  adaptive polar codes converge to RM codes.

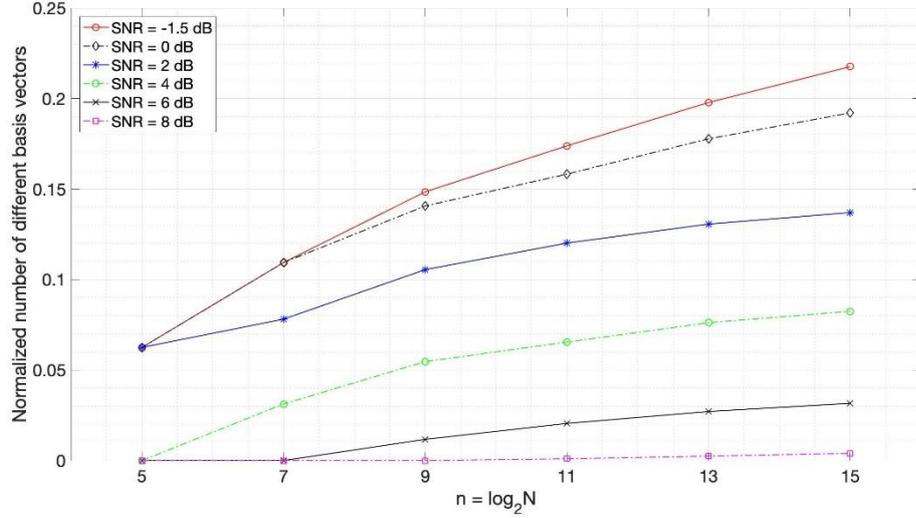
From Figure 4.9, one observes that the SNR value at which the two codes become the same increases with  $n$ . Similar to the definition of  $D(n, R, \epsilon)$  given in Section 4.2, we define  $D'(n, R, \text{SNR})$  as the “number of different rows between the generator matrices  $G_{RM}$  and  $G_P(\text{SNR})$  of rate- $R$ ,  $(2^n, K)$  RM codes and adaptive polar codes designed for an AWGN channel at a given SNR”. We then plot  $D'(n, 0.5, \text{SNR})$  versus  $n$  in Figure 4.10, for design SNRs chosen as 0, 2, 4 and 8 dB.

As  $n$  increases, the number  $D'(n, 0.5, \text{SNR})$  of different rows in Figure 4.10 seems to increase almost linearly on logarithmic scale; which is expected since the total number of basis vectors  $K = 2^{n-1}$  also grows exponentially with increasing  $n$ .



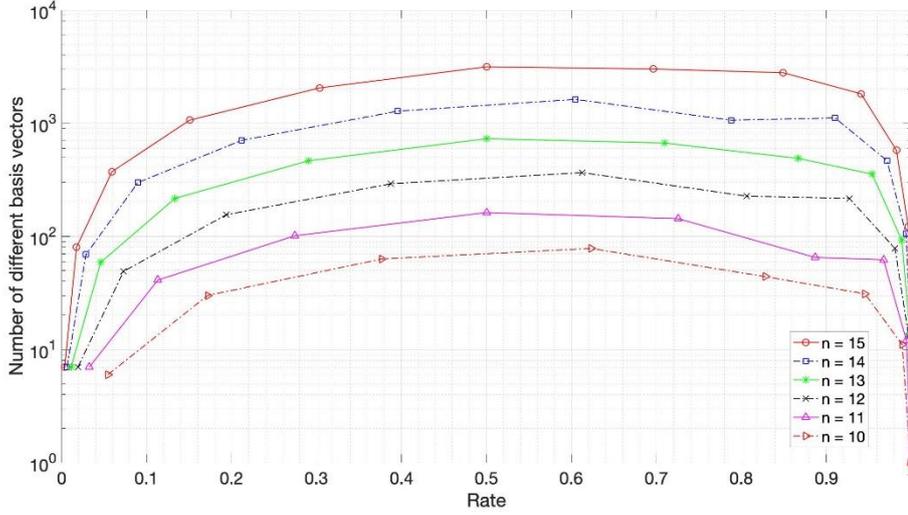
**Figure 4.10.** Number of different generator matrix rows  $D'(n, 0.5, SNR)$  of the  $(2^n, 2^{n-1})$  RM and adaptive polar codes designed for an AWGN at that SNR, versus  $n$ , for various design SNRs.

Due to the exponential increase observed in Figure 4.10, dividing  $D'(n, 0.5, SNR)$  by  $K$ , we obtain the normalized difference  $D'(n, 0.5, SNR)/K$  and plot in Figure 4.11. The vertical axis of Figure 4.11 is not logarithmic but linear. One notices that for  $SNR \geq 0$  dB and  $n \leq 15$ , the normalized ratio of differing rows between the generator matrices  $G_{RM}$  and  $G_P(SNR)$  of rate-0.5 codes remains less than 20%; in other words,  $D'(n, 0.5, SNR)/K \leq 0.20$  for  $SNR \geq 0$  dB. Examining the shape of each curve in Figure 4.11, one may also conjecture that as  $n$  grows further for a fixed value of the design SNR, the normalized number of different rows between two generator matrices approaches an asymptote. In addition to that,  $D'(n, 0.5, SNR)/K$  curves (given in Figure 4.11) and  $D(n, 0.5, \epsilon)/K$  curves (given in Figure 4.5) seem to behave quite similarly with respect to  $n$ .



**Figure 4.11.** Normalized number of different rows  $D'(n, 0.5, SNR)/K$ , between the generator matrices of the  $(2^n, 2^{n-1})$  RM and adaptive polar codes versus  $n$ , for various polar code design SNRs.

In the remaining part of this section, our objective is to examine the variation of  $D'(n, R, SNR)$  with respect to the code rate  $R = K/N$ . So, we start by fixing the SNR to 0 dB and observing the number of different rows,  $D'(n, R, 0)$  between the generator matrices of RM and polar codes designed for an AWGN channel with the SNR = 0 dB, at different code rates. Again, while designing the two codes, we primarily change the order  $r$  of the RM code and set  $(N, K)$  parameters of the polar code accordingly. In Figure 4.12, we plot  $D'(n, R, 0)$  versus rate  $R$ , for  $10 \leq n \leq 15$ ; where the vertical axis is scaled logarithmically.  $D'$  versus  $R$  curves of Figure 4.12 given for polar codes designed over an AWGN channel with SNR = 0 dB, look quite similar to  $D$  versus  $R$  curves in Figure 4.6, computed for polar codes designed over a BEC( $\epsilon$ ) with  $\epsilon = 0.1$ .



**Figure 4.12.** Number of differing rows  $D'(n, R, 0)$  between the generator matrices of the adaptive polar codes (designed for AWGN at  $SNR = 0$  dB) and RM codes, versus rate  $R$ .

One observes some properties from Figure 4.12, very similar to those in Figure 4.6; i.e.,

- i. For  $n_1 < n_2$ ,  $D'(n_1, R, 0) < D'(n_2, R, 0)$ . This is quite reasonable because the number of basis vectors  $K = 2^n R$  also increases as  $n$  increases, as discussed for the BEC case.
- ii. All curves are concave functions of the rate  $R$ . Denoting  $R_{0,n}$  as the rate where  $D'(n, R, 0)$  is maximized for each  $n$ , we observe  $R_{0,n}$  to be in the range  $(0.5, 0.65)$ . Each curve increases rapidly for  $0 < R < 0.2$  and have a relatively slower decrease for larger values of  $R$ . One observes a rapid decrease for  $R > 0.96$  and  $n \geq 12$ . This concavity can be reasoned similarly using (4.3) in Section 4.2.

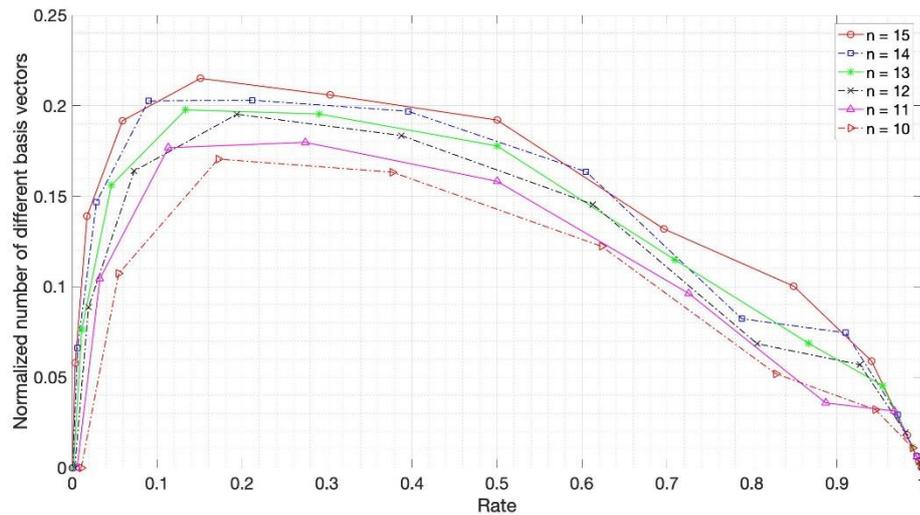
We then normalize  $D'(n, R, 0)$  shown in Figure 4.12 by dividing with the respective number of information bits  $K$ . Normalized difference  $D'(n, R, 0)/K$  is plotted in Figure 4.13, which is scaled linearly in the vertical axis. As in the BEC case plotted

for BEC(0.1), we observe a concave shape for the curves of  $D'(n, R, 0)/K$ , with  $R_{0,n}$  values at lower rates.

Similarly to the BEC case, we finally compute the normalized numbers  $D'(n, R, \text{SNR})/K$  versus  $R$ , considering different SNRs ranging from -5 to 7.5 dB with 2.5 dB steps, and  $10 \leq n \leq 15$ . Results are plotted in Figure 4.14. The corresponding channel capacities are tabulated in Table 4.2.

**Table 4.2.** Channel capacities with respect to given SNR values for the AWGN channel

SNR (dB)	-5.0	-2.5	0.0	2.5	5.0	7.5
Channel capacity $C = 0.5 \log_2(1 + \text{SNR})$	0.2	0.3	0.5	0.7	1.0	1.4



**Figure 4.13.** Normalized number  $D'(n, R, 0)/K$  of differing rows between the generator matrices of the adaptive polar codes designed for AWGN channel with  $\text{SNR} = 0$  dB and RM codes, versus the code rate  $R$ .

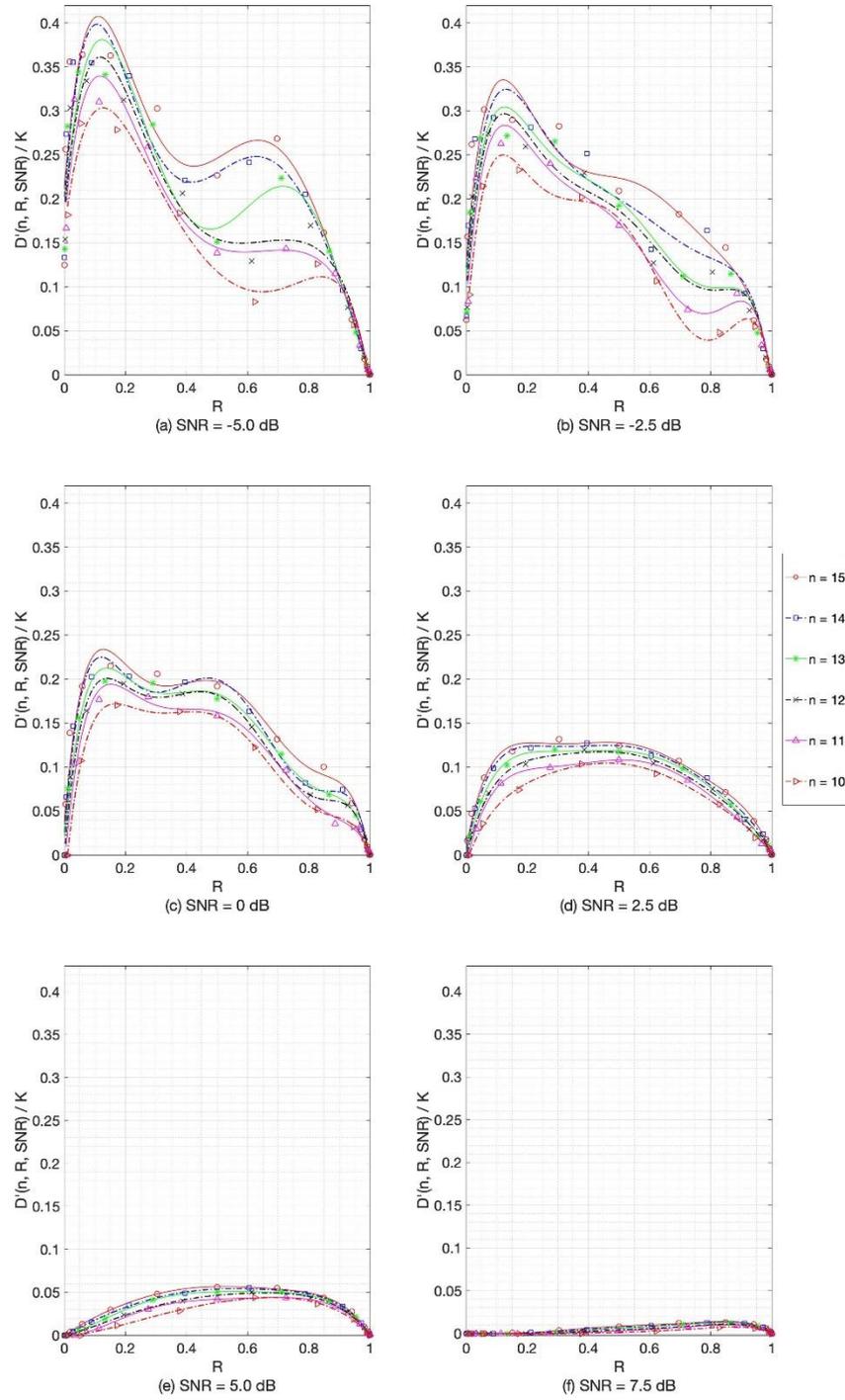
One may notice from Figure 4.14 that, the normalized ratio of differing rows between the generator matrices of RM codes and polar codes designed for AWGN at any rate

remains less than 40%; i.e.,  $D'(n, R, \text{SNR})/K \leq 0.40$  for  $10 \leq n \leq 15$  and  $\text{SNR} \geq -5$  dB; however, this ratio is more than the normalized difference ratio  $D(n, R, \epsilon)/K \leq 0.28$  given in Figure 4.8 for  $10 \leq n \leq 15$  and  $\epsilon \leq 0.5$ .

Comparing the six design-SNR values of Figure 4.14, additional observations are:

- i.* For  $n, R$  fixed, and  $\text{SNR}_1 < \text{SNR}_2$ ,  $[D(n, R, \text{SNR}_1)/K] > D(n, R, \text{SNR}_2)/K$ : This is again expected result [Mondelli, 2016, Proposition 6.1] as in BEC case since polar codes converge to RM codes as SNR increases. For all  $n \leq 15$ , the maximum of  $D'(n, R, \text{SNR})/K$  is found to be 37% at  $\text{SNR} = -5$  dB, which decreases to 30% at  $\text{SNR} = -2.5$  dB, 22% at  $\text{SNR} = 0$  dB, 13% at  $\text{SNR} = 2.5$  dB, 5.5% at  $\text{SNR} = 5$  dB, and 1.5% at  $\text{SNR} = 7.5$  dB.
- ii.* For small values of SNR,  $D'(n, R, \text{SNR})/K$  curves seem to be piecewise concave functions of the rate  $R$ , where there is a point  $\bar{R}_{\text{SNR},n} < C$  for each  $n$  and SNR such that  $D'(n, R, \text{SNR})/K$  is maximized. For smaller values of SNR, such concave-like behavior is only observed at meaningful transmission rates less than capacity; and at higher rates, a second peak starts to develop, as becomes especially apparent at  $\text{SNR} = -5$  dB. We still continue to denote  $\bar{R}_{\text{SNR},n} < C$ ; i.e., the first peak, as the rate value for which  $D'(n, R, \text{SNR})/K$  is maximum for meaningful rates  $R < C$ .
- iii.* As the design SNR increases, we observe that  $\bar{R}_{\text{SNR},n}$  increases; i.e., the peaks of  $D'(n, R, \text{SNR})/K$  versus  $R$  curves shift from left to right and start to occur at higher rates. Specifically, for  $\text{SNR} = -5, -2.5$  and  $0$  dB, one can see that the maximum values of  $D'(n, R, \text{SNR})/K$  occur for code rates in the  $(0.1, 0.3)$  range; while for  $\text{SNR} = 2.5$  dB,  $\bar{R}_{\text{SNR},n}$  is within the interval  $(0.3, 0.5)$ ; for  $\text{SNR} = 5$  dB, it is in the range  $(0.5, 0.6)$  and finally  $\bar{R}_{\epsilon,n}$  becomes more than 0.8 for  $\text{SNR} = 7.5$  dB.

In short, for  $\text{SNR}_1 < \text{SNR}_2$ , one concludes that  $\bar{R}_{\text{SNR}_1,n} < \bar{R}_{\text{SNR}_2,n}$ , where  $\bar{R}_{\text{SNR},n}$  is the rate value at which  $D'(n, R, \text{SNR})/K$  is maximum.



**Figure 4.14.** Normalized number  $D'(n, R, SNR)/K$  of differing rows between the generator matrices of the adaptive polar codes designed for AWGN channel and RM codes, versus the code rate  $R$ , at various SNRs.

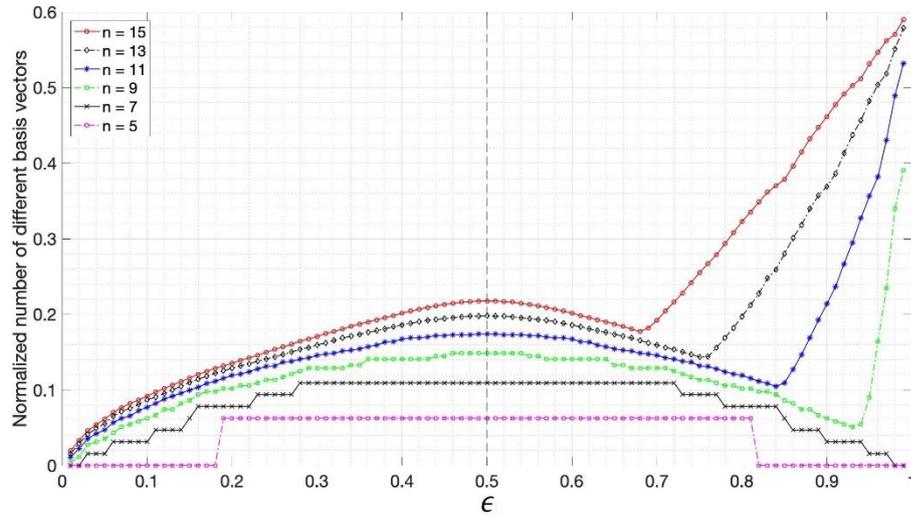
#### 4.4. Further Notes on the Convergence with respect to Channel Conditions

In the last part of this chapter, we present our further observations on the variation of  $D(n, 0.5, \epsilon)/K$  and  $D'(n, R, \text{SNR})/K$  (i.e., the normalized number of different vectors between generator matrices of the RM and polar codes) with respect to channel conditions.

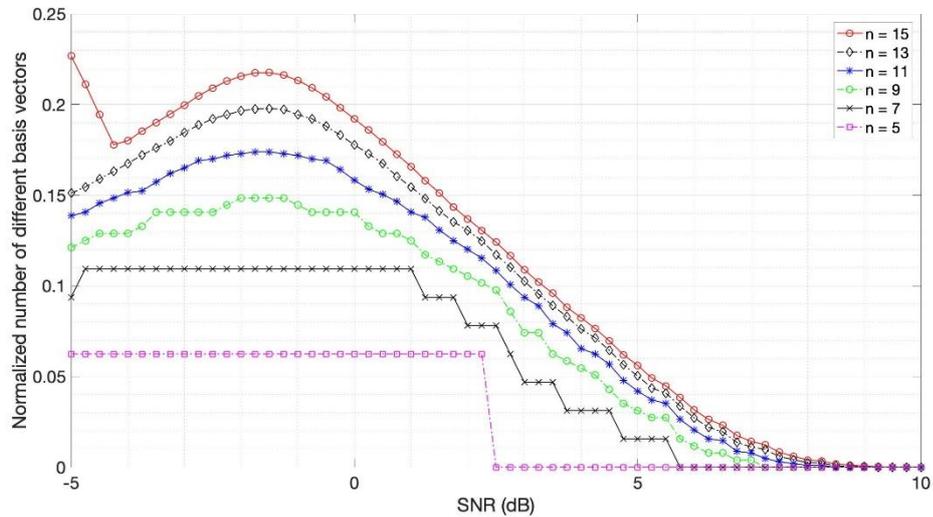
Due to aforementioned observations and Proposition 6.1 [Mondelli et al, 2014]; one could expect that the opposite of the proposition is also valid. In other words, as the channel gets worse, one may presume that the number of different rows between the generator matrices of the two codes would continuously increase. However, this is not the case, since the number of different vectors are limited by the minimum of  $K$  and  $N - K$  as explained by (4.2). Additionally, the maximum-weight vectors of  $G = F^{\otimes n}$  are also the ones with the smallest Bhattacharyya parameters, which have to be included in both generator matrices. Starting from the best channel conditions, say  $\epsilon = 0$ , where  $G_P = G_{RM}$ , the number of different vectors  $D(n, R, \epsilon)$  starts to increase as  $\epsilon$  increases up to a point, at which a high-weight basis vector previously discarded by the polar code rule may turn back into  $G_P$ , resulting in a decrease of  $D(n, R, \epsilon)$ .

Numerical calculations also show that as  $\epsilon$  increases, a decrease in  $D(n, 0.5, \epsilon)/K$  starts for  $\epsilon > 0.5$  as observed in Figure 4.15. This creates a piecewise concavity, but then turns into a surprisingly steep increase for higher values of  $\epsilon$  and  $n \geq 9$ . Concavity is lost because of this increase.

One also finds similar concavity with respect to the design SNR of the polar code. In Figure 4.16, we plot the normalized difference values  $D'(n, 0.5, \text{SNR})/K$  versus SNR and observe that although there is no significant change at smaller codelengths; for  $n \geq 11$ , the difference between polar and RM generator matrices starts to decrease as the SNR decreases beyond -1.5 dB. In addition, one can notice the steep increase of Figure 4.15 slightly in Figure 4.16 for  $n = 15$ .

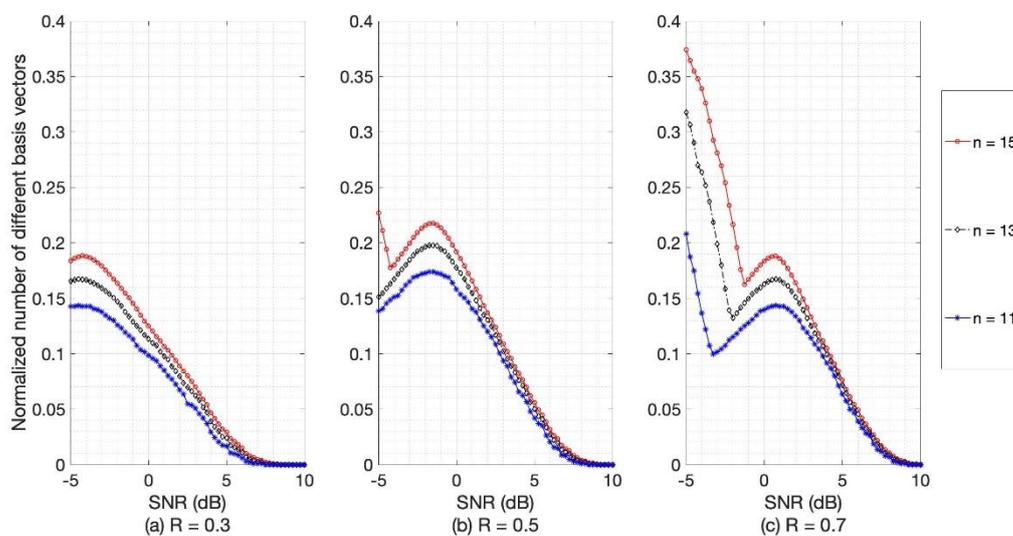


**Figure 4.15.** Normalized number of different rows  $D(n, 0.5, \epsilon)/K$ , between the generator matrices of the  $(2^n, 2^{n-1})$  RM and adaptive polar codes designed for that  $\epsilon$ , versus the design  $\epsilon$



**Figure 4.16.** Normalized number of different rows  $D'(n, 0.5, SNR)/K$ , between the generator matrices of the  $(2^n, 2^{n-1})$  RM and adaptive polar codes designed for that SNR, versus the design SNR of the adaptive polar code.

Lastly, in Figure 4.17, we compare  $D'(n, 0.5, SNR)/K$  of Figure 4.16 with normalized difference curves at rates 0.3 and 0.7; and observe that similar concavity starts for SNR values less than -4.25 dB and 0.75 dB, respectively. Furthermore, one can notice the steep increase of Figure 4.15 in part (c) of Figure 4.17, for all values of the codelength; i.e.,  $n = 11, 13$  and  $15$ .



**Figure 4.17.** Normalized number of different rows  $D(n, R, SNR)/K$ , between the generator matrices of the RM and adaptive polar codes designed for that  $SNR$ , versus the design  $SNR$  of the adaptive polar code at rates  $R = 0.3, 0.5$  and  $0.7$

## CHAPTER 5

### DISCUSSION AND CONCLUSION

In this thesis, two independent concepts related to the polar codes are investigated. First is the application of polar codes in block fading channels. For this purpose, we use the hierarchical polar coding scheme proposed in 2014 by Si, Köylüoğlu and Vishwanath (which we denote as the “SKV-code”) that uses multiple polar coding phases [Si et al., 2014]. The proposed scheme utilizes a BEC model for the fading blocks, in addition to some BSCs; and it has hierarchical stages in both the encoder and decoder. It is stated [Si et al., 2014] that since this scheme models a fading AWGN channel with a BPSK modulation and CSI only at the decoder, the results are relevant to practical cases in wireless communication.

Motivated by this asserted practicality, we compare the BER performance of the SKV-code to that of Arıkan’s original polar coding scheme under a fading channel model. We use a fixed code length of  $N = 256$ , number of blocks  $B = 256$  and only two fading states in our simulations. The two fading states yield a “good” channel with lower transition probability, and a “bad” channel with higher transition probability of the BSCs. We use different values for these probabilities, along with varying erasure probability of BEC that is used to model the block fading channels. We also use proper offsets to adjust the code rates such that  $R = \alpha C$ , where  $0 < \alpha < 1$  for practical reasons.

The SKV encoding and decoding algorithms are more complicated in the sense that they require more than one phase of encoding and decoding while Arıkan’s polar codes require only one phase; and the SKV decoder has the extra requirement of “knowing the channel states” at the decoder. Due to this tradeoff of increased complexity, we have expected to obtain a better BER performance when the proposed

scheme, specifically tailored for the fading channel, is used. In the light of the results given in Chapter 3; at small values of the probability of being in the degraded state, we observe that the performance curves of SKV-codes are always better than those of polar codes, which are constructed at the overall code rate of the two-state fading binary symmetric channels. However, as the probability of being in the degraded state,  $q_1$ , increases, Arıkan's original polar coding scheme constructed with respect to either the bad or good channel's crossover probability performs much better than the SKV-code. Moreover, as  $q_1$  increases, we observe the SKV-codes are converging to the uncoded BER performance due to the errors made by the BEC decoder.

We note that we have assumed only two fading states for simplicity and we have not been able to simulate the SKV-code for larger code lengths such as 2048 due to computational latency, which is a drawback of our MATLAB programs. On the other hand, the authors [Si et al., 2014] claim that their scheme works better than Arıkan's polar code when the code length  $N$  and block number  $B$  are large, at rates very close to the channel capacity. Furthermore, the performance of SKV-code may be improved when different decoders other than SC such as BP, SCL or BPL are used. In fact, Arıkan's polar code may also benefit from such decoders, especially the list decoders, in fading channels as Trifonov points out [Trifonov, 2015] stating that the decoding error probability relies strongly on the minimum distance of the codes used and in order to obtain a good performance, sequential or list decoding should be used. However, since the complexity is relatively high for the SKV-code, we believe that using a more complex decoder to improve performance would not be appropriate due to practical reasons.

The second topic we have discussed in this thesis is the convergence of polar codes to RM codes. Noting that the two codes are close relatives, we are interested in observing the convergence behavior in the light of the related proposition by Mondelli [Mondelli, 2016] and finding the convergence points of the two codes with respect to the polar code design parameters of the related BEC and AWGN channels. For this purpose, we compare the rows of the generator matrix of the adaptive polar code designed for the

specific channel, with that of the RM code of same code length and rate. We firstly compute the erasure probabilities of BECs, and SNRs of AWGN channels, where the adaptive polar code becomes the same as the RM code.

We add some behavioral observations relating to this convergence: For both of the channels, we may conjecture that as code length  $N = 2^n$  of rate-0.5 codes grows for a fixed value of the design erasure probability or SNR, the normalized number of different rows between two generator matrices also grows and approaches an asymptote. For code lengths smaller than  $2^{15}$ , the normalized ratio of differing rows remains less than 22% and 20% for the BEC and AWGN channels, respectively.

We express the difference  $D(n, R, \epsilon)$  for BECs or  $D'(n, R, \text{SNR})$  for AWGN channels as a function of the code length, rate and design parameter, which can be either the erasure probability  $\epsilon$  or the SNR, depending on the channel of interest. We also perform the comparison over the normalized number of different rows, normalized with respect to the total number of generator matrix rows, to observe the difference percentage. Using this notation, the previous paragraph summarizes the behaviors of  $D(n, 0.5, \epsilon)/K$  and  $D'(n, 0.5, \text{SNR})/K$  versus  $n$ .

As for the behavior versus  $R$ , we observe that for the polar codes designed over BECs, difference percentages  $D(n, R, \epsilon)/K$  are piecewise concave functions of the rate  $R$ , whose first maxima occur for rates less than the channel capacity. These normalized difference curves remain less than 28% when the erasure probability is smaller than 0.5 and code lengths are between  $2^{10}$  and  $2^{15}$ . Similar to the BEC case, one observes similar concave shapes in the normalized difference values for the polar codes designed over AWGNs. The normalized difference curves remain less than 40%, which is more than that of the BEC case. In both cases,  $D(n, R, \epsilon)$  and  $D'(n, R, \text{SNR})$  curves with respect to rate  $R$  start and end at 0: Denoting  $F^{\otimes n}$  of (2.26) as the mother-generator matrix  $G$  whose rows are chosen according to the RM-rule for  $G_{RM}$  and the polar rule for  $G_P$ ; at rate  $1/N$ , all-one vector of  $G$  is the first basis vector of both codes, and is chosen as the first element of  $G_{RM}$  and  $G_P$ . This makes  $D(n, R, \epsilon)$

and  $D'(n, R, \text{SNR})$  both equal to zero. At rate  $N/N = 1$ , the generator matrices of the two codes are equal to the mother generator matrix, i.e.;  $G = G_{RM} = G_P$ , which again makes the number of different basis vectors between the two codes for both the BEC and AWGN channel equal to zero; i.e.,  $D(n, R, \epsilon) = D'(n, R, \text{SNR}) = 0$ .

While investigating the conditions under which polar codes approach RM codes with respect to the channel parameters, we have come across a behavior that needs to be clarified as the channel gets worse; i.e., as the design parameter  $\epsilon$  becomes higher or the SNR gets smaller. Due to aforementioned observations and Proposition 6.1 [Mondelli et al, 2014]; as the channel gets worse, one may expect that the number of different rows between the generator matrices of the two codes would continuously increase, assuming that the opposite of the proposition is also valid. However, as explained by (4.2), the number of different vectors are limited by the minimum of  $K$  and  $N - K$ , which implies that a continuous increase is not possible. Additionally, the maximum-weight vectors of the mother-generator matrix  $G = F^{\otimes n}$  are also the ones with the smallest Bhattacharyya parameters, which have to be included in both generator matrices. Starting from the best channel conditions where the generator matrix  $G_P$  of the polar code is equal to  $G_{RM}$ , the number of different vectors starts to increase as  $\epsilon$  increases (or SNR decreases) up to a point, at which a high-weight basis vector previously discarded by the polar code rule may turn back into  $G_P$ , resulting in a decrease in the number of different rows. We give these results in more detail in Part 4.4.

Regarding the code lengths used in Chapter 4, we could not increase it further from  $2^{15}$  due to the lack of computational memory. Nonetheless, we do not think that similar computation for higher code lengths would make a significant contribution to the present results of this study.

Last but not least, as Mondelli states that this convergence can be generalized to any BMS channel, we have also computed the numerical data for BSCs at different transition probabilities and rates. However, we could not observe the convergence as

directly as we did for BEC and AWGN channels. We believe that a more comprehensive investigation is needed for convergence on the BSC and will be putting this issue under the scope of our future work.



## REFERENCES

- Abbe, E., Shpilka, A., & Wigderson, A. (2015). Reed-Muller Codes for Random Erasures and Errors. *IEEE Transactions on Information Theory*, 61(10), 5229–5252, doi: 10.1109/TIT.2015.2462817
- Akdoğan, S. C. (2018). *A Study on the Set Choice of Multiple Factor Graph Belief Propagation Decoders for Polar Codes*. Middle East Technical University.
- Arıkan, E. (2008). A performance comparison of polar codes and reed-muller codes. *IEEE Communications Letters*, 12(6), 447–449, doi: 10.1109/LCOMM.2008.080017
- Arıkan, E. (2009). Channel polarization: A method for constructing capacity-achieving codes. *IEEE International Symposium on Information Theory - Proceedings*, 55(7), 1173–1177, doi: 10.1109/ISIT.2008.4595172
- Arıkan, E. (2010a). A survey of Reed-Muller codes from polar coding perspective. *IEEE Information Theory Workshop 2010 (ITW 2010)*, (4), 1–5, doi: 10.1109/ITWKSPS.2010.5503223
- Arıkan, E. (2010b). Polar codes: A pipelined implementation. *4th International Symposium on Broadband Communication (ISBC 2010)*, 2–4.
- Arıkan, E. (2012). Polar Coding: ISIT 2012 Tutorial.
- Arıkan, E. (2015). On the Origin of Polar Coding. *IEEE Journal on Selected Areas in Communications*, 34(2), 209–223, doi: 10.1109/JSAC.2015.2504300
- Bahl, L. R., Cocke, J., Jelinek, F., & Raviv, J. (1974). Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate. *Symposium A Quarterly Journal in Modern Foreign Literatures*, (January 1972), 284–287.

- Berrou, C., Glavieux, A., & Thitimajshima, P. (1993). Claude Berrou, Alain Glavieux. *1993 IEEE International Conference on Communications*, (1).
- Blahut, R. E. (1984). *Theory and Practice of Error Control Codes*. Owego, NY: Addison-Wesley Publishing Company Inc., IBM Corporation.
- Bose, R. C., & Ray-Chaudhuri, D. K. (1960). On a class of error correcting binary group codes. *Information and Control*, 3(1), 68–79, doi: 10.1016/S0019-9958(60)90287-4
- Boutros, J. J., & Biglieri, E. (2013). Polarization of quasi-static fading channels. *IEEE International Symposium on Information Theory - Proceedings*, (October), 769–773, doi: 10.1109/ISIT.2013.6620330
- Bravo-Santos, A. (2013). Polar codes for the rayleigh fading channel. *IEEE Communications Letters*, 17(12), 2352–2355, doi: 10.1109/LCOMM.2013.111113.132103
- Chen, K., Niu, K., & Lin, J. (2012). List successive cancellation decoding of polar codes. *Electronic Letters*, 48(9), doi: 10.1109/TCOMM.2013.070213.120789
- Costello, D. J., & Forney, G. D. (2007). Channel coding: The road to channel capacity. *Proceedings of the IEEE*, 95(6), 1150–1177, doi: 10.1109/jproc.2007.895188
- Cover, T. M., Thomas, J. A. (2006). *Elements of Information Theory*. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Doan, N., Hashemi, S. A., Mondelli, M., & Gross, W. J. (2018). On the Decoding of Polar Codes on Permuted Factor Graphs, 0–5. Retrieved from <http://arxiv.org/abs/1806.11195>
- Dumer, I. (2004). Recursive Decoding and Its Performance for Low-Rate Reed-Muller Codes, 50(5), 811–823.

- Dumer, I. (2006). Soft-Decision Decoding of Reed – Muller Codes: A Simplified Algorithm, *52*(3), 954–963.
- Dumer, I. (2017). Recursive decoding of Reed-Muller codes, doi: 10.1109/isit.2000.866353
- Dumer, I., & Shabunov, K. (2017). Soft-decision decoding of reed-muller codes: Recursive lists. *IEEE Transactions on Information Theory*, *52*(3), 1260–1266, doi: 10.1109/TIT.2005.864443
- Elias, P. (1955). Coding for Noisy Channels. *Proceedings of the IRE*, *43*, 356–356.
- Elkelesh, A., Cammerer, S., Ebada, M., & Ten Brink, S. (2017). Mitigating clipping effects on error floors under belief propagation decoding of polar codes. *Proceedings of the International Symposium on Wireless Communication Systems, 2017–August*, 384–389, doi: 10.1109/ISWCS.2017.8108145
- Elkelesh, A., Ebada, M., Cammerer, S., & Ten Brink, S. (2018a). Belief propagation decoding of polar codes on permuted factor graphs. *IEEE Wireless Communications and Networking Conference, WCNC, 2018–April*, 1–6, doi: 10.1109/WCNC.2018.8377158
- Elkelesh, A., Ebada, M., Cammerer, S., & Ten Brink, S. (2018b). Belief Propagation List Decoding of Polar Codes. *IEEE Communications Letters*, *22*(8), 1536–1539, doi: 10.1109/LCOMM.2018.2850772
- Forney, G. D. (1970). Convolutional Codes I: Algebraic Structure. *IEEE Transactions on Information Theory*, *IT-16*, 720–738.
- Forney, G. D. (2001). Codes on Graphs: Normal Realizations. *IEEE Transactions on Information Theory*, *47*(2), 520–548.
- Gallager, R. G. (1962). Low-Density Parity-Check Codes. *IRE Transactions on Information Theory*, 21–28, doi: 10.1002/0471667846.ch9

- Golay, M. J. E. (1949). Notes on Digital Coding. *Bell System Technical Journal*, 27, 418.
- Hamming, R. W. (1950). Error detecting and error correcting codes.pdf. *Bell System Technical Journal*, 29(2).
- Hashemi, S. A., Balatsoukas-Stimming, A., Giard, P., Thibeault, C., & Gross, W. J. (2016). Partitioned successive-cancellation list decoding of polar codes. *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (L1), 957–960, doi: 10.1109/ICASSP.2016.7471817
- Hashemi, S. A., Doan, N., Mondelli, M., & Gross, W. J. (2018). Decoding Reed-Muller and Polar Codes by Successive Factor Graph Permutations, 1–5, doi: 10.1109/istc.2018.8625281
- Hassani, S. H., Condo, C., Gross, W. J., Urbanke, R. L., Mondelli, M., & Hashemi, S. A. (2018). Decoder Partitioning: Towards Practical List Decoding of Polar Codes. *IEEE Transactions on Communications*, 66(9), 3749–3759, doi: 10.1109/tcomm.2018.2832207
- Hocquenghem, A. (1959). Codes correcteurs d'erreurs. *Chiffres (paris)*, 2(147-156):116.
- Hussami, N., Korada, S. B., & Urbanke, R. L. (2009). Performance of Polar Codes for Channel and Source Coding. *ISIT*, 1488–1492.
- Korada, S. B. (2009). *Polar Codes for Channel and Source Coding*. Ecole Polytechnique Federale de Lausanne. Retrieved from [https://infoscience.epfl.ch/record/138655/files/EPFL\\_TH4461.pdf](https://infoscience.epfl.ch/record/138655/files/EPFL_TH4461.pdf)
- Kudekar, S., Kumar, S., Mondelli, M., Pfister, H. D., Sasoğlu, E., & Urbanke, R. L. (2017). Reed-muller codes achieve capacity on erasure channels. *IEEE Transactions on Information Theory*, 63(7), 4298–4316, doi: 10.1109/TIT.2017.2673829

- Lee, W. C. Y. (2004). *Mobile Communications Engineering*. McGraw-Hill.
- Li, B., Shen, H., & Tse, D. (2014). A RM-Polar Codes, 1–2. Retrieved from <http://arxiv.org/abs/1407.5483>.
- Li, H., & Yuan, J. (2013). A practical construction method for polar codes in AWGN channels. *IEEE 2013 Tencon - Spring, TENCONSpring 2013 - Conference Proceedings*, 223–226, doi: 10.1109/TENCONSpring.2013.6584444
- Liu, S., Hong, Y., & Viterbo, E. (2017). Polar Codes for Block Fading Channels. *2017 IEEE Wireless Communications and Networking Conference Workshops, WCNCW 2017*, 1–6, doi: 10.1109/WCNCW.2017.7919041
- MacKay, D. J. C., & Neal, R. M. (2002). Near Shannon limit performance of low density parity check codes. *Electronics Letters*, 32(18), 1645, doi: 10.1049/el:19961141
- Mondelli, M. (2016). *From Polar to Reed-Muller Codes: Unified Scaling, Non-standard Channels, and a Proven Conjecture*. Ecole Polytechnique Federalde de Lausanne.
- Mondelli, M., Hassani, S. H., & Urbanke, R. L. (2014). From polar to reed-muller codes: A technique to improve the finite-length performance. *IEEE Transactions on Communications*, 62(9), 3084–3091, doi: 10.1109/TCOMM.2014.2345069
- Muller, D. E. (1954). Application of Boolean algebra to switching circuit design and to error detection. *Transactions of the I.R.E. Professional Group on Electronic Computers, EC-3(3)*, 6–12, doi: 10.1109/irepgelc.1954.6499441
- Murata, T., & Ochiai, H. (2017). On design of CRC codes for polar codes with successive cancellation list decoding. *IEEE International Symposium on Information Theory - Proceedings*, 1868–1872, doi: 10.1109/ISIT.2017.8006853

Niu, K., & Chen, K. (2012). CRC-aided decoding of polar codes. *IEEE Communications Letters*, 16(10), 1668–1671, doi: 10.1109/LCOMM.2012.090312.121501

Özgür, Ü. (2009). *A Performance Comparison of Polar Codes with Convolutional Turbo Codes*. Bilkent University.

Rappaport, T. S. 1996. *Wireless Communications Principles and Practice*. Upper Saddle River, NJ: Prentice-Hall Inc.

Reed, I. S. (1954). A Class of Multiple-Error-Correcting Codes and the Decoding Scheme. Retrieved from <http://ir.obihiro.ac.jp/dspace/handle/10322/3933>

Reed, I. S., & Solomon, G. (1960). Polynomial codes over certain finite fields. *Journal of the society for industrial and applied mathematics*, 8(2): 300-304.

Richardson, T., & Urbanke, R. L. (2008). *Modern Coding Theory*. New York: Cambridge University Press.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27, 379–423.

Si, H., Köylüoğlu, O. O., & Vishwanath, S. (2014). Polar coding for fading channels: Binary and exponential channel cases. *IEEE Transactions on Communications*, 62(8), 2638–2650, doi: 10.1109/TCOMM.2014.2345399

Tal, I., & Vardy, A. (2013). How to construct polar codes. *IEEE Transactions on Information Theory*, 59(10), 6562–6582, doi: 10.1109/TIT.2013.2272694

Tal, I., & Vardy, A. (2015). List Decoding of Polar Codes. *IEEE Transactions on Information Theory*, 61(5), 2213–2226, doi: 10.1109/TIT.2015.2410251

- Tanner, R. M. (1981). A recursive approach to low complexity codes. *IEEE Transactions on Information Theory*, IT-27, 533–547.
- Trifonov, P. (2015). Design of polar codes for Rayleigh fading channel. *Proceedings of the International Symposium on Wireless Communication Systems, 2016–April(2)*, 331–335, doi: 10.1109/ISWCS.2015.7454357
- Tse, D., & Viswanath, P. (2005). *Fundamentals of Wireless Communication*. New York: Cambridge University Press.
- Vangala, H., Viterbo, E., & Hong, Y. (2014). Permuted successive cancellation decoder for polar codes. *2014 International Symposium on Information Theory and Its Applications*, (C), 438–442.
- Viterbi, A. J. (1967). Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm. *IEEE Transactions on Information Theory*, 13(2), 260–269.
- Wozencraft, J. M. (1957). Sequential Decoding for Reliable Communication. *National IRE Convention Record*, 5(2), 11–25.
- Yuan, B., & Parhi, K. K. (2014). Early stopping criteria for energy-efficient low-latency belief-propagation polar code decoders. *IEEE Transactions on Signal Processing*, 62(24), 6496–6506, doi: 10.1109/TSP.2014.2366712



## APPENDICES

### A. Proofs of Some of the Equations Used

#### A.1. Fading Binary Symmetric Channel Modeling

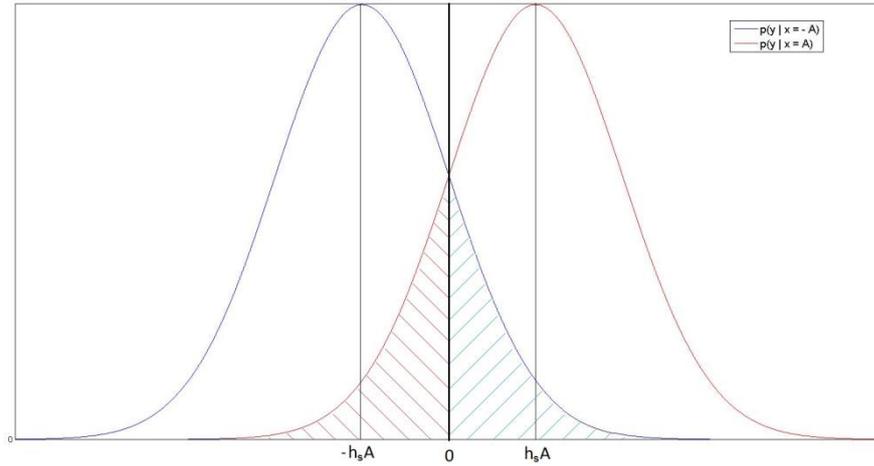
A fading AWGN channel with BPSK modulation is assumed and its correspondence to a fading binary symmetric channel as given in Equation (3.2) is proved here. A fading AWGN channel can be written as

$$Y = h_s X + Z$$

$$X = \begin{cases} A, & \text{with probability } 1/2 \\ -A, & \text{with probability } 1/2 \end{cases}$$

$Z \sim \mathcal{N}(0, \sigma)$ : zero mean Gaussian noise is assumed.

Since the probability distribution function (pdf) of the Gaussian random variable  $Z$  is  $p_Z(z) = \frac{1}{\sqrt{2\pi}\sigma} e^{-z^2/2\sigma^2}$ , then the received output  $Y$  given that  $X$  is sent has a distribution  $p_Y(y|x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-\mu_x)^2/2\sigma^2}$  where  $\mu_x \in \{h_s A, -h_s A\}$  when  $x \in \{A, -A\}$  are sent respectively. The corresponding pdfs are shown in Figure A.1.



**Figure A.1.** Probability distribution function of  $p_Y(y|x)$

Error occurs when either  $\{y < 0|x = A\}$  or  $\{y > 0|x = -A\}$  events occur. Since the areas under the probability distribution functions of these events (shaded with red and blue lines in Figure A.1, respectively) are the same, the probability of error in both cases are the same. As a result, assuming that these event occur equally likely, it is sufficient to show the error probability of just one of the events alone:

$$\begin{aligned} P_e &= \frac{1}{2}P\{y < 0|x = A\} + \frac{1}{2}P\{y > 0|x = -A\} \\ &= P\{y < 0|x = A\} \end{aligned} \quad (\text{A.1})$$

Thus, we will show the probability of error of the prior case. The channel gain of this channel can be denoted as  $h_s = h$  for simplicity.

For the event  $\{y < 0|x = A\}$ , the probability of error is simply the area as shown in red dashes:

$$P_e = \int_{-\infty}^0 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-hA)^2/2\sigma^2} dx \quad (\text{A.2})$$

With a change of variable  $u = \frac{x-hA}{\sigma}$ , we have

$$P_e = \int_{-\infty}^{hA/\sigma} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \quad (\text{A.3})$$

Denoting the cumulative normal function of the Gaussian distribution as  $\Phi(x)$

$$\Phi(x) = P\{\mathcal{N}(0,1) < x\} = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \quad (\text{A.4})$$

Then, comparing (A.3) and (A.4), we have  $P_e = 1 - \Phi(hA/\sigma)$ . Since the signal-to-noise ratio is  $SNR = \frac{P_x}{P_z} = \frac{A^2}{\sigma^2}$ , then the probability of error of this scheme is simply  $P_e = 1 - \Phi(h\sqrt{SNR})$ .

In the modeled fading BSC, a two-state channel (with channel gains  $h_1$  and  $h_2$ ) is assumed, which happen with probabilities  $q_1$  and  $q_2$ , and the noise at each state is distributed with  $\text{Bern}(p_1)$  and  $\text{Bern}(p_2)$  respectively. In other words, error occurs with probabilities  $p_s$  where  $s \in \{1,2\}$ . As a result, we can model each fading channel as  $W_s \triangleq \text{BSC}(p_s)$  with probability  $q_s$  where

$$p_s \triangleq P_e = 1 - \Phi(h_s \sqrt{SNR}). \quad (\text{A.5})$$

We would like to add the following, without changing the result of the proof: Recall the *Q-function*  $Q(x) = P\{\mathcal{N}(0,1) > x\}$ , then  $P_e$  can also be written as  $Q(h\sqrt{SNR})$  and thus  $p_s = Q(h_s\sqrt{SNR})$ . ■

## A.2. Minimum Degree of the Polynomial $Z_n^{(i)}(x)$

We prove by induction that  $Z_n^{(i)}(x)$  is a polynomial with minimum degree equal to  $2^{w_H(\mathbf{b}^{(i)})}$  where  $w_H(\cdot)$  is the Hamming weight and  $\mathbf{b}^{(i)}$  is the binary expansion of  $i - 1$  over  $n = \log_2 N$  bits, i.e.,  $\mathbf{b}^{(i)} = b_1^{(i)} b_2^{(i)} \dots b_n^{(i)}$ ,  $b_1^{(i)}$  being the most significant bit and  $b_n^{(i)}$  being the least significant bit.

- Proof for  $n = 1$ :

For  $n = 1$ , the binary expansion over 1 bit can be either 0 or 1 (We can represent only 2 indices with 1 bit):

$$\mathbf{b}^{(i)} = b_1^{(i)} = \begin{cases} 0, & i = 1 \\ 1, & i = 2 \end{cases}, \quad (\text{A.6})$$

Then, the corresponding Bhattacharyya parameters are

$$Z_1^{(i)}(x) = f_{b_1^{(i)}} = \begin{cases} f_0(x) = 2x - x^2, & i = 1 \\ f_1(x) = x^2, & i = 2 \end{cases}, \quad (\text{A.7})$$

As can be seen, when binary expansion of the indices over 1 bit is performed, the minimum degree of  $Z_1^{(1)}(x)$  is equal to  $2^{w_H(b_1^{(1)})} = 2^0 = 1$  and that of  $Z_1^{(2)}(x)$  is equal to  $2^{w_H(b_1^{(2)})} = 2^1 = 2$ .

- Assume that above claim holds for  $n = k$ :

For  $n = k$ , the binary expansion of any index  $i$  is  $\mathbf{b}^{(i)} = b_1^{(i)} b_2^{(i)} \dots b_k^{(i)}$  and it is assumed that the polynomial  $Z_k^{(i)}(x) = f_{b_1^{(i)}} \circ f_{b_2^{(i)}} \circ \dots \circ f_{b_k^{(i)}}(x)$  having a minimum degree of  $2^{w_H(\mathbf{b}^{(i)})}$  is true. Observe that the (maximum) degree of this polynomial is  $2^k$  since both of  $f_0(x)$  and  $f_1(x)$  have degree of 2 and they are repeated by substitution  $k - \text{times}$ .

- Show that above claim holds for  $n = k + 1$ :

For  $n = k + 1$ , the binary expansion of any index  $i$  is  $\mathbf{b}'^{(i)} = b_1^{(i)} b_2^{(i)} \dots b_k^{(i)} b_{k+1}^{(i)}$  and the Bhattacharyya parameter is  $Z_{k+1}^{(i)}(x) = f_{b_1^{(i)}} \circ f_{b_2^{(i)}} \circ \dots \circ f_{b_k^{(i)}} \circ f_{b_{k+1}^{(i)}}(x)$ . Recall that in the binary expansion of an integer over  $m$  bits,  $b_1^{(i)}$  is the most significant bit and  $b_m^{(i)}$  is the least significant bit; and if one more bit is used in the expansion ( $m \leftarrow m + 1$  bits) the newly added bit becomes the new most significant bit (i.e., it is inserted from the left of the previous  $m - \text{bit}$  expansion). As a result, we already know that the right of  $f_{b_1^{(i)}}$  is a polynomial with a minimum degree of  $2^{w_H(\mathbf{b}^{(i)})}$  from the previous step of the proof and we can write it as

$$Z_k^{(i)}(x) = ax^{2^k} + bx^{2^{k-1}} + \dots + cx^{2^{w_H(\mathbf{b}^{(i)})}} \quad (\text{A.8})$$

The new most significant bit  $b_1^{(i)}$  can either be 0 or 1, and depending on its value,  $Z_{k+1}^{(i)}(x)$  can be calculated by substituting  $Z_k^{(i)}(x)$  in  $f_0(x)$  and  $f_1(x)$ , respectively:

$$Z_{k+1}^{(i)}(x) = \begin{cases} f_0(Z_k^{(i)}(x)), & b_1^{(i)} = 0 \\ f_1(Z_k^{(i)}(x)), & b_1^{(i)} = 1 \end{cases} \quad (\text{A.9})$$

For the first case, we need to show that the minimum degree is still  $2^{w_H(\mathbf{b}^{(i)})}$  since  $b_1^{(i)} = 0$  implies  $w_H(\mathbf{b}'^{(i)}) = w_H(\mathbf{b}^{(i)})$ . It is easy to see that due to the first term of  $f_0(x)$ , which is  $2x$ , this is in fact true. For the second case, we need to show that the minimum degree has increased by 1 since  $b_1^{(i)} = 1$  implies  $w_H(\mathbf{b}'^{(i)}) = w_H(\mathbf{b}^{(i)}) + 1$ . As  $f_1(x) = x^2$ , the last term of  $(Z_k^{(i)}(x))^2$  becomes  $c^2 x^{2*2^{w_H(\mathbf{b}^{(i)})}}$  which results in minimum degree of  $Z_{k+1}^{(i)}(x)$  being  $2^{w_H(\mathbf{b}^{(i)})+1} = 2^{w_H(\mathbf{b}'^{(i)})}$ .

■