TECHNICAL UNIVERSITY OF CRETE
ELECTRICAL AND COMPUTER ENGINEERING DEPARTMENT

Polar Coding for the Binary Erasure Channel

by

Eleni Antoniou

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DIPLOMA DEGREE OF

ELECTRICAL AND COMPUTER ENGINEERING

May 2018

THESIS COMMITTEE

Associate Professor Georgios N. Karystinos, *Thesis Supervisor*
Professor Athanasios P. Liavas
Associate Professor Aggelos Bletsas

Abstract

In this thesis, we present the basic principles of polar coding for the binary erasure channel. We explain the process of channel polarization and demonstrate that polar coding is capacity achieving. Then, we present efficient techniques for the coding process, two different implementations of the successive decoder, and an efficient method for the construction of the code (i.e., the selection of the virtual channels that will carry the useful information). Finally, we show how polar coding can be used for the degraded broadcast channel to transmit public data to two receivers and private data to one of them.

Acknowledgements

I would like to thank my friends and my family for supporting me during my studies and especially my sister Eleftheria. Also, I would like to thank my supervisor, Professor Georgios Karystinos, for his guidance and assistance throughout these months.

Contents

List of Figures

Chapter 1

Introduction

Polar codes are the first explicitly proven codes with implementable complexity that can achieve Shannon capacity, due to Professor Erdal Arikan, whose breakthrough in 2009 created a huge interest in academia [2]. Polar codes can be used in communication links that are vulnerable to errors due to random noise, interference etc. that disrupt the original data stream at the receiving end. Channel coding basically uses a set of algorithmic operations on the original data stream at the transmitter and another set on the received data stream at the receiver, to correct these errors. These operations at the transmitter and receiver are respectively denoted as encoding and decoding operations [3].

The main goal of channel coding is to develop high-performance channel codes that eliminate the errors in communications as much as possible. The real challenge here is to accomplish this in low complexity that allows practical implementation. The complexity of a code determines everything, e.g., how much power it consumes, how much memory it needs, all of which that determine whether a code is good for any practical use or not [3].

The polar-coding transformation involves two key operations called channel combining and channel splitting. At the encoder, channel combining assigns combinations of bits to specific channels. The channel splitting that follows performs an implicit transformation operation, translating these bit combinations into decoder-ready vectors. The decoding operation at the receiver tries to estimate the original input bits by using a successive-cancellation decoding technique [3].

Channel combining, channel splitting, and successive-cancellation decoding essentially convert a block of bits into a polarized bit stream at the receiver. In this way, a received bit and its associated channel end up being either a "good channel" or "bad channel." It has been mathematically proven that, as the size of the bit block increases, the received bit stream polarizes in a way that the number of "good channels" approaches Shannon capacity. This phenomenon is what gives polar codes their name and makes them the first explicitly proven capacity-achieving channel codes [3].

In channel polarization, that we are going to explain in Chapter 2, we construct a set of N channels out of N independent copies of a given B-DMC W, by combining and splitting those copies. Two examples of binary-input symmetric-output memoryless channels are the binary symmetric channel (BSC) and the binary erasure channel (BEC).

1.1 Symmetric Capacity

Given a binary-discrete memoryless channel (B-DMC) W with input alphabet $\mathcal{X} = \{0, 1\}$ and output alphabet $\mathcal{Y}$, we define the symmetric capacity as

$$I(W) = \sum_{y \in Y} \sum_{x \in X} \frac{1}{2} W(y|x) \log_2 \frac{W(y|x)}{\frac{1}{2}W(y|0) + \frac{1}{2}W(y|1)}. \tag{1.1}$$

Symmetric capacity is the mutual information between the input and the output of the channel, when the input distribution is uniform ($P(X = 0) = P(X = 1) = \frac{1}{2}$). Consequently, when the channel is symmetric (where the optimal input distribution is the uniform), its Shannon capacity is equal to its symmetric capacity.

1.2 Bhattacharyya Parameter

In addition to the symmetric capacity above, we introduce the Bhattacharyya parameter of a B-DMC W, which is used as measure of reliability of the channel, as it constitutes an upper bound on the probability of maximum-likelihood (ML) decision error when W is used only once to transmit a 0 or 1 [1].

$$Z(W) = \sum_{y \in Y} \sqrt{W(y|0)W(y|1)}. \tag{1.2}$$

The following are true for any B-DMC W channel, showing the relation between the symmetric capacity and the Bhattacharyya parameter

$$I(W) \geqslant \log \frac{2}{1 + Z(W)}, \tag{1.3}$$

$$I(W) \leq \sqrt{1 + Z(W)^2}. \tag{1.4}$$

1.3 Binary Symmetric Channel (BSC)

The binary symmetric channel, shown in Fig. 1.1, has input one of two binary digits $(0, 1)$ and output the input digit, either correctly with probability $1 - p$ or inverted with p. The symmetric capacity of BSC is

$$I(W) = 1 + p \log_2(p) + (1 - p) \log_2(1 - p). \tag{1.5}$$

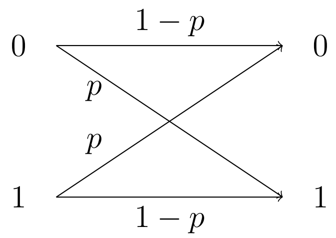


Figure 1.1: BSC(p)

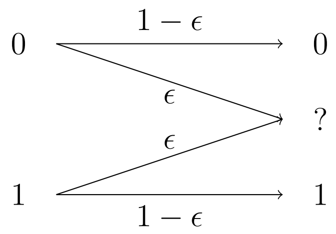


Figure 1.2: BEC(ϵ)

1.4 Binary Erasure Channel (BEC)

The binary erasure channel has input one of two binary digits $(0, 1)$ and output either the same digit with probability $1-\epsilon$ or a message that the bit was not received with probability ϵ, as shown in Fig. 1.2. In BEC there is no incorrect information. The symmetric capacity of BEC is

$$I(W) = 1 - \epsilon. \tag{1.6}$$

Chapter 2

Polarization for Binary-input Channels

Channel polarization is a method for constructing code sequences that achieve the symmetric capacity $I(W)$ of any binary-input discrete memoryless channel W. The symmetric capacity is the highest rate achievable, subject to using the input letters of the channel with equal probability. In this chapter we will explain the construction of the Polar codes, proposed by E. Arikan [1].

2.1 Basic Polarization

Channel polarization is an operation by which one manufactures a new set of N channels $\{W_N^{(i)} : 1 \leq i \leq N\}$ out of N independent copies of a given discrete memoryless channel W, that show a polarization effect in the sense that, as N becomes large, the symmetric capacity terms tend towards 0 or 1 for all but a vanishing fraction of indices i [1].

To achieve this effect, we combine two independent copies of binary-input channel W_1 to obtain the channel $W_2 : \mathcal{X}^2 \to \mathcal{Y}^2$ with the transition probabilities (2.1) and (2.2) [1]. Fig. 2.1 shows how to construct channel W_2. We set x_1 to be the result of the XOR between u_1 and u_2 and x_2 to be equal to u_2. By combining a pair of copies of W, we construct a new composite channel denoted as W_2 with two bits as input and two bits as output. After this transformation, the first channel degrades and the second upgrades in terms of symmetric capacity.

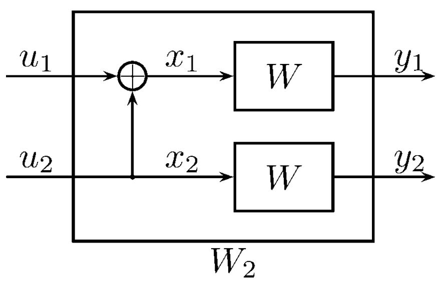


Figure 2.1: The channel W_2.

We define the channel $W_2 : \mathcal{X}^2 \to \mathcal{Y}^2$, shown in Fig. 2.1 as

$$\begin{aligned} W_2(y_1, y_2|u_1, u_2) &= W_2(y_1, y_2|x_1, x_2) \\ &= W(y_1|x_1)W(y_2|x_2) = W(y_1|u_1 \oplus u_2)W(y_2|u_2), \end{aligned}$$

with the following transition probabilities for each channel

$$\begin{aligned} {W_2}^{(1)}(y_1, y_2|u_1) &= \sum_{u_2} p(y_1, y_2|u_1, u_2)p(u_2|u_1) \\ &= \sum_{u_2} p(y_1, y_2|u_1, u_2)p(u_2) \\ &= \frac{1}{2}p(y_1, y_2|u_1, u_2 = 0) + \frac{1}{2}p(y_1, y_2|u_1, u_2 = 1) \\ &= \frac{1}{2}p(y_1, y_2|x_1, x_2 = 0) + \frac{1}{2}p(y_1, y_2|x_1, x_2 = 1) \\ &= \frac{1}{2}p(y_1|x_1)p(y_2|x_2 = 0) + \frac{1}{2}p(y_1|x_1)p(y_2|x_2 = 1) \\ &= \frac{1}{2}\sum_{u_2} p(y_1|x_1)p(y_2|x_2) \\ &= \frac{1}{2}\sum_{u_2} p(y_1|u_1 \oplus u_2)p(y_2|u_2), \end{aligned} \tag{2.1}$$

$$\begin{aligned} {W_2}^{(2)}(y_1, y_2, u_1|u_2) &= p(y_1, y_2|u_1, u_2)p(u_2|u_1) \\ &= p(y_1, y_2|u_1, u_2)p(u_2) \\ &= \frac{1}{2}p(y_1, y_2|u_1, u_2) \\ &= \frac{1}{2}p(y_1, y_2|x_1, x_2) \\ &= \frac{1}{2}p(y_1|x_1)p(y_2|x_2) \\ &= \frac{1}{2}p(y_1|u_1 \oplus u_2)p(y_2|u_2), \end{aligned} \tag{2.2}$$

for all $u_1, u_2 \in \mathcal{X}, y_1, y_2 \in \mathcal{Y}$, with u_1, u_2 independent.

In the Appendix, we calculate the previous transition probabilities one-by-one, for both BSC and BEC, and then we use them in (1.1) to compute the symmetric capacity of the ${W_2}^{(1)}$ and ${W_2}^{(2)}$ channels.
After the calculations, we observe that, for the BSC,

$$I(u_1; y_1, y_2) < 1 - H(p) \text{ and } I(u_2; y_1, y_2, u_1) > 1 - H(p)$$

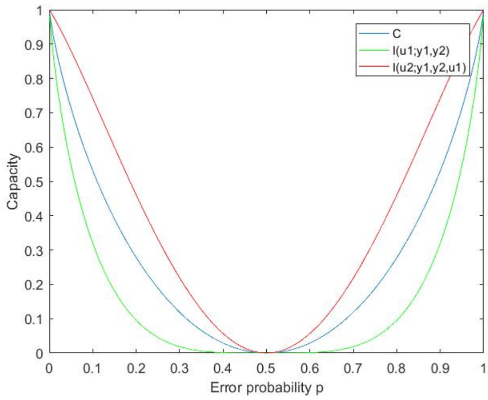
and, for the BEC,

$$I(u_1; y_1, y_2) < 1 - \varepsilon \text{ and } I(u_2; y_1, y_2, u_1) > 1 - \varepsilon.$$

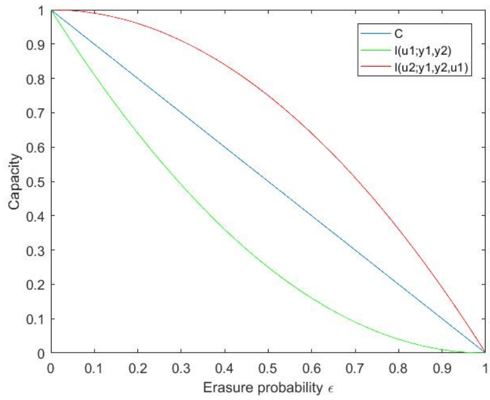
All the above are shown in Fig.2.2a and Fig.2.2b.

The two bit channels created after the basic polarization step are defined as upgraded and degrated channels. We observe that for both BSC and BEC, the degraded channel has smaller capacity than the original channel and the upgraded bigger (Fig. 2.2a and Fig. 2.2b). This property is mathematically expressed as

$$I({W_2}^{(1)}) \leqslant I(W) \leqslant I({W_2}^{(2)}). \tag{2.3}$$



(a) BSC Capacity.



(b) BEC Capacity.

Figure 2.2: Symmetric capacity of BSC and BEC, before and after the basic polarization step.

It is important to mention that the single step channel transformation preserves the symmetric capacity:

$$I({W_2}^{(1)}) + I({W_2}^{(2)}) = 2I(W_1). \tag{2.4}$$

Proof.

$$\begin{aligned}
I({W_2}^{(1)}) + I({W_2}^{(2)}) &= I(u_1; y_1, y_2) + I(u_2; y_1, y_2, u_1) \\
&= I(u_1; y_1, y_2) + I(u_2; y_1, y_2|u_1) + I(u_2; u_1), \text{ according to chain rule} \\
&= I(u_1; y_1, y_2) + I(u_2; y_1, y_2|u_1),\ u_2, u_1 \text{ are i.i.d. so } I(u_2; u_1) = 0, \\
&= I(u_1, u_2; y_1, y_2) \\
&= I(x_1, x_2; y_1, y_2) \\
&= I(x_1; y_1) + I(x_2; y_2) = \boldsymbol{2I(W_1)}. \qquad \square
\end{aligned}$$

Additionally, we can also prove that:

$$I({W_4}^{(1)}) + I({W_4}^{(2)}) + I({W_4}^{(3)}) + I({W_4}^{(4)}) = 4I(W_1) \tag{2.5}$$

Proof.

$$I({W_4}^{(1)}) + I({W_4}^{(2)}) + I({W_4}^{(3)}) + I({W_4}^{(4)}) =$$
$$= I(u_1; y_1, y_2, y_3, y_4) + I(u_2; y_1, y_2, y_3, y_4, u_1) +$$
$$+ I(u_3; y_1, y_2, y_3, y_4, u_1, u_2) + I(u_4; y_1, y_2, y_3, y_4, u_1, u_2, u_3)$$

(according to the chain rule),

$$= I(u_1; y_1, y_2, y_3, y_4) + I(u_2; y_1, y_2, y_3, y_4|u_1) + I(u_2; u_1) +$$
$$+ I(u_3; y_1, y_2, y_3, y_4, u_2|u_1) + I(u_3; u_1) + I(u_4; y_1, y_2, y_3, y_4, u_2, u_3|u_1) + I(u_4; u_1)$$

(u_4, u_3, u_2, u_1 are i.i.d. so $I(u_2; u_1) = I(u_3; u_1) = I(u_4; u_1) = 0$)

$$= I(u_1; y_1, y_2, y_3, y_4) + I(u_2; y_1, y_2, y_3, y_4) + I(u_3; y_1, y_2, y_3, y_4|u_2) +$$
$$+ I(u_3; u_2) + I(u_4; y_1, y_2, y_3, y_4, u_2|u_2) + I(u_4; u_2)$$

(u_4, u_3, u_2 are i.i.d. so $I(u_3; u_2) = I(u_4; u_2) = 0$)

$$= I(u_1; y_1, y_2, y_3, y_4) + I(u_2; y_1, y_2, y_3, y_4) + I(u_3; y_1, y_2, y_3, y_4) +$$
$$+ I(u_4; y_1, y_2, y_3, y_4|u_3) + I(u_4; u_3)$$

(u_4, u_3 are i.i.d. so $I(u_2; u_3) = 0$)

$$\begin{aligned}
&= I(u_1; y_1, y_2, y_3, y_4) + I(u_2; y_1, y_2, y_3, y_4) + I(u_3; y_1, y_2, y_3, y_4) + I(u_4; y_1, y_2, y_3, y_4) \\
&= I(u_1, u_2, u_3, u_4; y_1, y_2, y_3, y_4) \\
&= I(x_1, x_2, x_3, x_4; y_1, y_2, y_3, y_4) \\
&= I(x_1; y_1) + I(x_2; y_2) + I(x_3; y_3) + I(x_4; y_4) = \boldsymbol{4I(W_1)}.
\end{aligned}$$ □

Finally, the above properties can be generalized for N bits:

$$\sum_{i=1}^{N} I(W_N{}^{(\mathrm{i})}) = NI(W) \tag{2.6}$$

Proof.

$$\begin{aligned}
\sum_{i=1}^{N} I(W_N{}^{(\mathrm{i})}) &= I(u_1; y_1^N) + I(u_2; y_1^N, u_1) + I(u_3; y_1^N, u_1, u_2) + ... + I(u_N; y_1^N, u_1^{N-1}) \\
&= I(u_1; y_1^N) + I(u_2; y_1^N|u_1) + ... + I(u_N; y_1^N|u_1^{N-1}) \\
&= I(u_1 ... u_N; y_1 ... y_N) = I(x_1 ... x_N; y_1, ... y_N) \\
&= I(x_1; y_1) + I(x_2; y_2) + I(x_3; y_3) + ... + I(x_N; y_N) = NI(W)
\end{aligned}$$

with $u_1, u_2, ..., u_n$ i.i.d.

For the BEC, the following two properties are true

$$I({W_2}^{(1)}) = I(W)^2, \tag{2.7}$$

$$I({W_2}^{(2)}) = 2I(W) - I(W)^2. \tag{2.8}$$

Proof.

According to (1.2),

$$\begin{aligned}
Z({W_2}^{(2)}) &= \sum_{y_1^2, u_1} \frac{1}{2} \sqrt{{W_2}^{(2)}(y_1, y_2, u_1|0)} \sqrt{{W_2}^{(2)}(y_1, y_2, u_1|1)} \\
&= \sum_{y_1^2, u_1} \frac{1}{2} \sqrt{W(y_1|u_1 \oplus 0) W(y_2|0)} \sqrt{W(y_1|u_1 \oplus 1) W(y_2|1)} \\
&= \sum_{y_1^2, u_1} \frac{1}{2} \sqrt{W(y_1|u_1) W(y_2|0)} \sqrt{W(y_1|u_1 \oplus 1) W(y_2|1)} \\
&= \sum_{y_2} \sqrt{W(y_2|0) W(y_2|1)} \sum_{u_1} \frac{1}{2} \sum_{y_1} \sqrt{W(y_1|u_1) W(y_1|u_1 \oplus 1)} \\
&= Z(W) \cdot 1 \cdot Z(W) \\
&= Z(W)^2
\end{aligned}$$

For any BEC(ε) W, $\boldsymbol{Z(W) = \varepsilon}$,

$$\begin{aligned}
Z(W) &= \sum_{y \epsilon \Upsilon} \sqrt{W(y|0) W(y|1)} \\
&= \sqrt{W(0|0) W(0|1)} + \sqrt{W(e|0) W(e|1)} + \sqrt{W(1|0) W(1|1)} \\
&= \sqrt{(1-\varepsilon) \cdot 0} + \sqrt{\varepsilon \cdot \varepsilon} + \sqrt{0 \cdot (1-\varepsilon)} \\
&= \boldsymbol{\varepsilon}
\end{aligned}$$

We also know that the symmetric capacity of BEC is

$$I(W) = 1 - \varepsilon.$$

We conclude that $\boldsymbol{Z(W) = 1 - I(W)}$ for BEC.

$$
\begin{aligned}
&Z({W_2}^{(2)}) = Z(W)^2 \Leftrightarrow \\
&\Leftrightarrow 1 - I({W_2}^{(2)}) = (1 - I(W))^2 \\
&\Leftrightarrow 1 - I({W_2}^{(2)}) = 1 + I(W)^2 - 2I(W) \\
&\Leftrightarrow \boldsymbol{I({W_2}^{(2)}) = 2I(W) - I(W)^2}
\end{aligned}
$$

considering that $2I(W) = I({W_2}^{(1)}) + I({W_2}^{(2)})$, we can also prove that

$\boldsymbol{I({W_2}^{(1)}) = I(W)^2}$. $\square$

In the following figures, we also show in Matlab that, when we implement the transition probabilities calculated at the Appendix, their properties are confirmed.

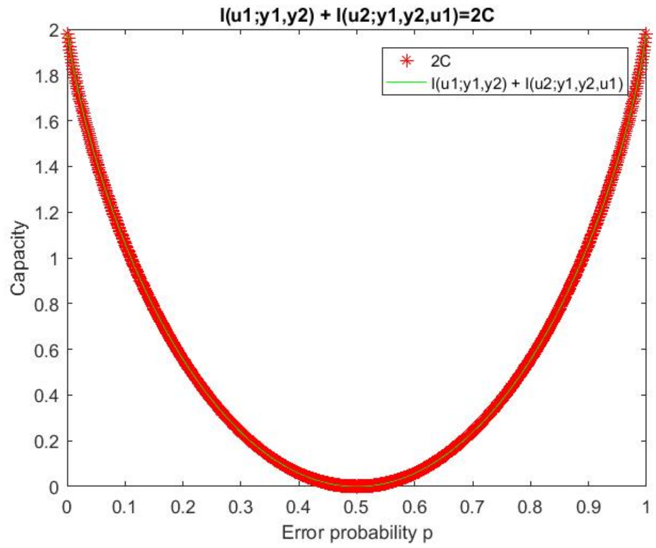


Figure 2.3: $I({W_2}^{(0)}) + I({W_2}^{(1)}) = 2C$ for BSC.

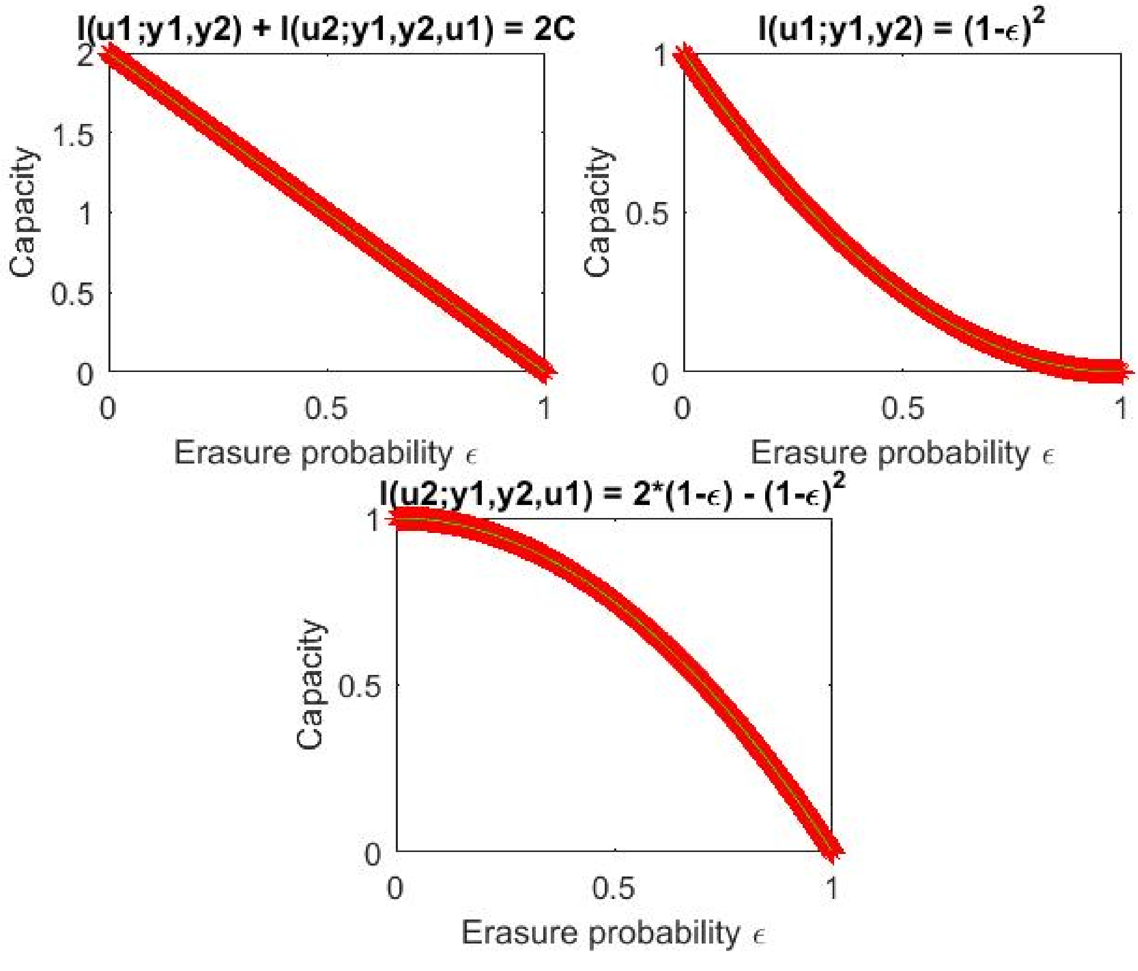


Figure 2.4: Properties of BEC.

We also computed $I(u_2; y_1, y_2)$ and $I(u_1; y_1, y_2, u_2)$ to see if the polarization step works in this case. In the figures we realized that those two capacities are exactly the same with the initial Capacity C, therefore the channel polarization step does not apply in this situation. Nevertheless, the capacity preserving property (2.4) is still true.

Proof.

$$I(u_2; y_1, y_2) = I(u_2; y_1|y_2) + I(u_2; y_2), \text{according to chain rule.}$$

But,

$$\begin{aligned} I(u_2; y_1|y_2) &= I(y_1; u_2|y_2) = I(y_1; y_2, u_2) - I(y_1; y_2) \\ &= I(y_1; u_2) + I(y_1; y_2|u_2) - I(y_1; y_2) \end{aligned}$$

We observe that $y_2 \to u_2 \to y_1$ and $u_2 \to x_1 \to y_1$ are Markov chains. From the first Markov chain we notice that y_1 is independent from y_2 when we know u_2, so $I(y_1; y_2|u_2) = 0$.

From the second chain, we notice that $I(u_2; x_1) \geq I(u_2; y_1)$, but $I(u_2; x_1) = 0$ because u_2 and x_1 are independent, so $I(u_2; y_1) = 0$.

In conclusion, $I(u_2; y_1|y_2) = 0 + 0 - I(y_1; y_2)$, but $I(u_2; y_1|y_2)$ cannot be negative, so $I(y_1; y_2) = 0$ which means that

$$I(u_2; y_1, y_2) = I(u_2; y_2) = I(x_2; y_2) = I(W_1), \tag{2.9}$$

$$I(u_1, u_2; y_1, y_2) = I(x_1, x_2; y_1, y_2) = I(x_1; y_1) + I(x_2; y_2) = 2I(W_1). \tag{2.10}$$

$I(u_1, u_2; y_1, y_2)$ can also be written as

$$I(u_1, u_2; y_1, y_2) = I(u_1; y_1, y_2|u_2) + I(u_2; y_1, y_2)$$

$I(u_1; y_1, y_2|u_2) = I(u_1; y_1, y_2, u_2) - I(u_1; u_2)$, ($u_1, u_2$ are independent which means $I(u_1; u_2) = 0$, so)

$$\Rightarrow I(u_1; y_1, y_2|u_2) = I(u_1; y_1, y_2, u_2). \tag{2.11}$$

From (2.9), (2.10), and (2.11), we can prove that

$$\begin{aligned} I(u_1, u_2; y_1, y_2) &= I(u_1; y_1, y_2|u_2) + I(u_2; y_1, y_2) \Rightarrow \\ 2I(W_1) &= I(u_1; y_1, y_2, u_2) + I(W_1) \Rightarrow \\ I(u_1; y_1, y_2, u_2) &= I(W_1). \end{aligned}$$

□

We can recursively calculate the symmetric capacities of the manufactured channels using the following formulas (2.12), (2.13) with $I(W) = 1 - \epsilon$ [1]. This whole process shows the polarization effect, meaning that, as N becomes large, the symmetric capacity tends to become either 0 or 1. We decide that the information bits are sent across the noiseless channels ($I(W)$ close to one) and the frozen (fixed) bits are sent across the noisy ones ($I(W)$ close to zero).
In the case of BEC(0.5), the symmetric capacity is 0.5 bits per channel use. After implementing (2.12) and (2.13) in Matlab, we observe that almost half of the channels are perfect and the other half are useless. This is the effect of channel polarization.

$$I({W_N}^{(\text{2i-1})}) = I({W_{N/2}}^{(\text{i})})^2 \tag{2.12}$$

$$I({W_N}^{(\text{2i})}) = 2I({W_{N/2}}^{(\text{i})}) - I({W_{N/2}}^{(\text{i})})^2 \tag{2.13}$$

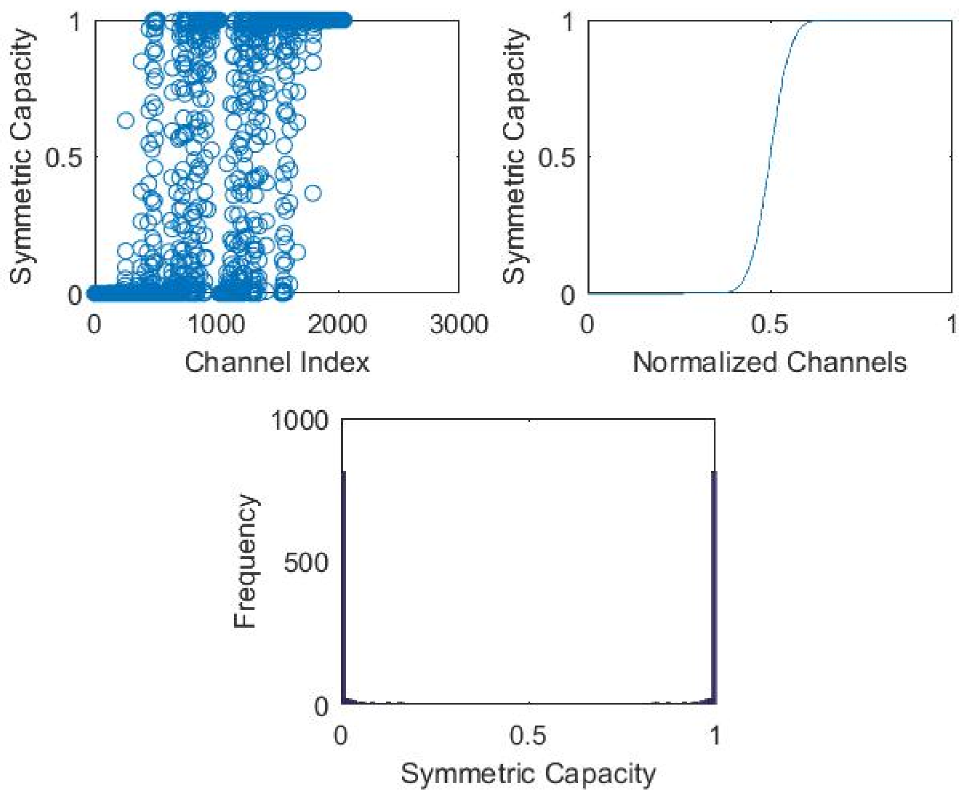


Figure 2.5: Channel polarization for a BEC with $\varepsilon = 0.5$ and $N = 2^{11}$.

In the case of BEC(0.2), the symmetric capacity is 0.8 bits per channel use. Using (2.12) and (2.13) in Matlab, we observe that almost the 80% of the channels are perfect and the 20% are useless.

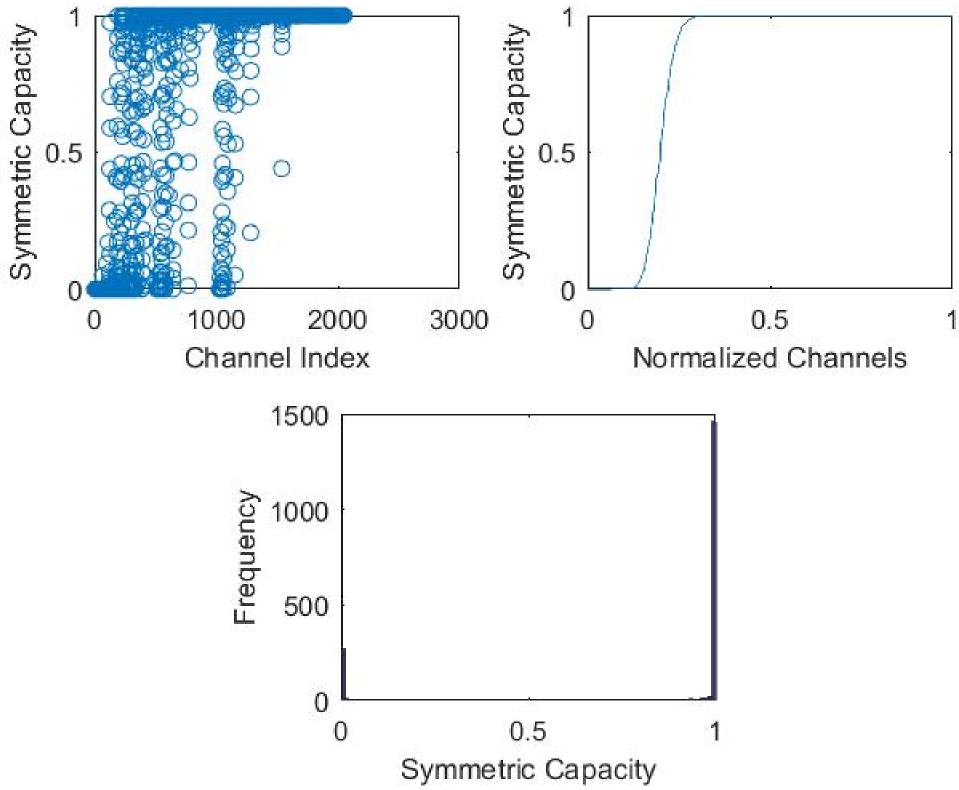


Figure 2.6: Channel polarization for a BEC with $\varepsilon = 0.2$ and $N = 2^{11}$.

2.2 Encoding

In Fig. 2.8, at the random Nth level of polarization we combine two copies of $W_{N/2}$. For an array of inputs $\{u_i\}_{i=1}^{N}$ we obtain a new one $\{v_i\}_{i=1}^{N}$, by replacing the odd indexed u's with the XOR of consecutive pairs, which means that the odd indexed u_{2i-1} will be equal to $u_i \oplus u_{i+1}$, for each $i = 1,\ldots, N/2$. The even indexed u's, on the other hand, remain as they are. The operator R_N is a permutation, known as the reverse shuffle operation, which simply separates the odd-indexed from the even-indexed elements [1]. Odd-indexed signals become input to the first copy of $W_{N/2}$ and even-indexed to the second one. The permutation essentially rotates cyclically the bit-indexes to the left. More specifically, the first input v_1 is expressed as $v_{00\ldots0}$, where the string of zeros has length $n = \log_2 N$. Accordingly, all the rest $N-1$ v's are expressed in the same binary way. Essentially, the permutation changes the order of this bit string cyclically, by right shifting by one the bit-indices. For example, for $n = 2$ (Fig. 2.7), the input $u_1^4 = (u_{00}, u_{01}, u_{10}, u_{11})$ is transformed to $v_1^4 = (u_{00}, u_{10}, u_{01}, u_{11})$, which means that $u_1^4 = (u_1, u_2, u_3, u_4)$ is converted to $v_1^4 = (u_1, u_3, u_2, u_4)$. The outputs of the two channels W_2 are y_1^4.
The encoding complexity, with the help of Master Theorem, is

$$\begin{aligned} T(N) &= \frac{N}{2} + \Theta(N) + 2T(\frac{N}{2}) \Rightarrow \\ T(N) &= \Theta(N \log_2 N). \end{aligned} \tag{2.14}$$

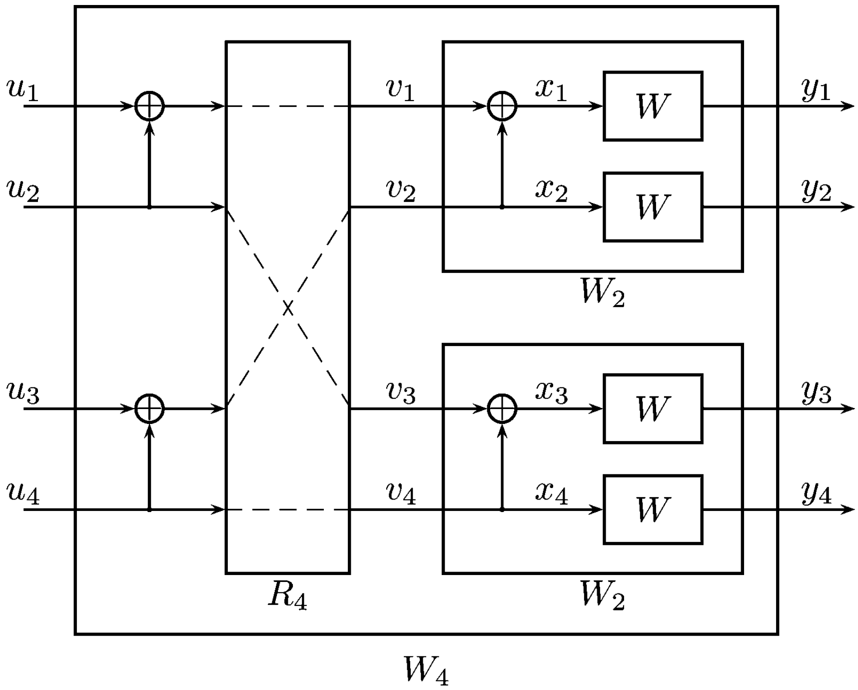


Figure 2.7: The channel W_4 and its relation to W_2 and W.

The generalized encoding scheme, where the recursive construction of W_N from two copies of $W_{N/2}$ is depicted.

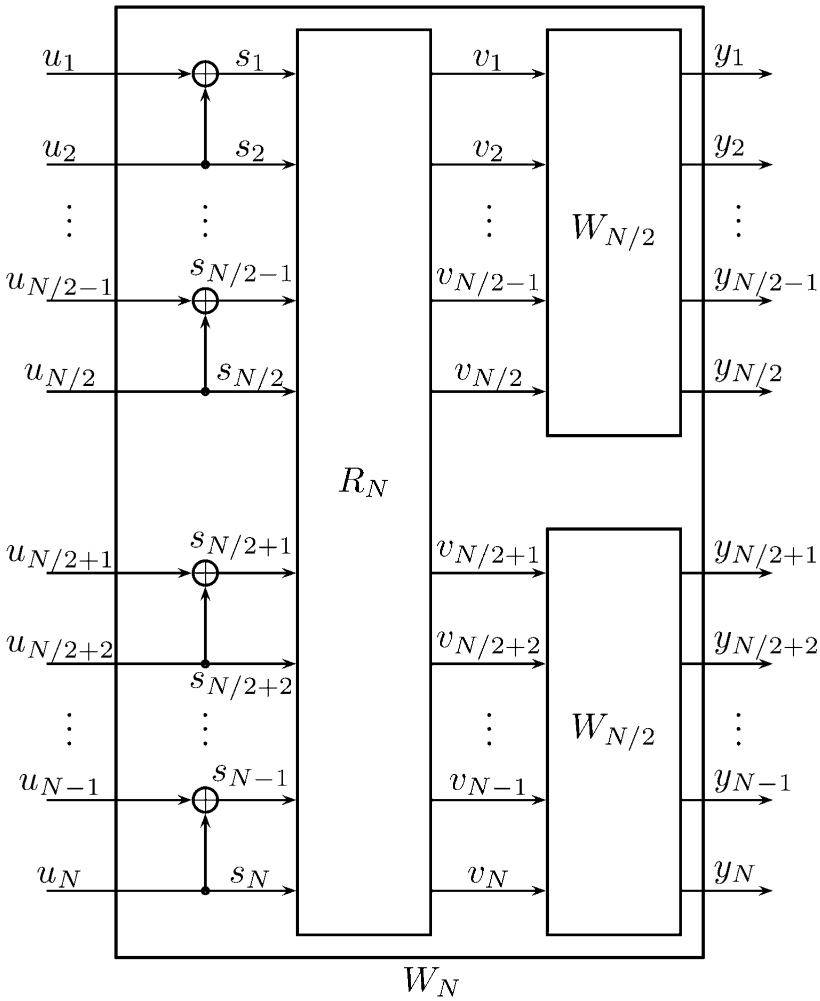


Figure 2.8: Recursive construction of W_N from two copies of $W_{N/2}$.

2.3 Successive Cancellation Decoding

The successive cancellation (SC) decoder, introduced in [1], decides with the rule of closest neighbor on the ith bit ($1 \leq i \leq N$) that is transmitted over $W_N^{(i)}$, by computing (2.15) and (2.16) for BEC and (2.17) and (2.18) for BSC.

More specifically, for the BEC with erasure probability ε, first we use the recursive formulas of the symmetric capacity (2.12), (2.13). After the recursive calculation, we assign the frozen bits (usually 0's) wherever the capacity has its lowest prices, so that we can use the good channels (those with high capacity) to transmit the information bits.

To decode successfully, we calculate the transition probabilities by using the efficient recursive formulas (2.15), (2.16). Each estimation is carried out by using the knowledge of frozen and previously estimated symbols. More specifically, if u_i is a frozen bit, then $\hat{u}_i$ takes the value of u_i, and, if u_i is an information bit, then the decoder computes the transition probabilities for BEC or the likelihood ratios (LR) for BSC.

We implemented two decoding approaches, a slow (complexity $O(N^2)$) and a fast (complexity $O(N \log_2 N)$). In the slow decoder, we store each calculation of (2.15) and (2.16) in a matrix and we estimate the original value of $\hat{u}$ by comparing $W_N^{(i)}(y_1^N, u_1^{i-1}|0)$ and $W_N^{(i)}(y_1^N, u_1^{i-1}|1)$. The fast, on the contrary, was generated by observing that several computed values are calculated more than one time, so we store them in a cell array of size $N \times (\log_2 N + 1)$ to avoid recalculation and to use the stored values when needed. In each cell, both probabilities $W_N^{(i)}(y_1^N, u_1^{i-1}|0)$ and $W_N^{(i)}(y_1^N, u_1^{i-1}|1)$, are stored so that we can compare them easily and use them for the calculations. Each cell is filled after $\Theta(1)$ calculations, which implies that the complexity of this decoding is $O(N \log_2 N)$. The calculation of a probability at length N for both decoders is reduced to the calculation of two probabilities. Accordingly, at length $N/2$, each calculation of a probability at length $N/2$ requires the calculation of two probabilities at length $N/4$ and so on. This recursion can be computed until it reaches the block length 1, at which point we calculate $W(y_i|0)$ and $W(y_i|1)$.

The following decoding scheme estimates sequentially every information symbol for BEC:

$$\hat{u}_i \triangleq \begin{cases} u_i & \text{, when } u_i \text{ is a frozen bit,} \\ \arg\max_{x \epsilon (0,1)} W_N^{(i)}(y_1^N, u_1^{i-1}|x) & \text{, otherwise,} \end{cases}$$

using the recursive formulas

$$W_{2N}^{(2i-1)}(y_1^{2N}, u_1^{2i-2}|u_{2i-1}) = \sum_{u_{2i}} \frac{1}{2} W_N^{(i)}(y_1^N, u_{1,o}^{2i-2} \oplus u_{1,e}^{2i-2}|u_{2i-1} \oplus u_{2i}) \cdot W_N^{(i)}(y_{N+1}^{2N}, u_{1,e}^{2i-2}|u_{2i}), \tag{2.15}$$

and

$$W_{2N}^{(2i)}(y_1^{2N}, u_1^{2i-1}|u_{2i}) = \frac{1}{2} W_N^{(i)}(y_1^N, u_{1,o}^{2i-2} \oplus u_{1,e}^{2i-2}|u_{2i-1} \oplus u_{2i}) \cdot W_N^{(i)}(y_{N+1}^{2N}, u_{1,e}^{2i-2}|u_{2i}). \tag{2.16}$$

For the BSC, we followed the same procedure as in BEC for the slow decoder. For the fast decoder, we also took advantage of the fact that some of the recursions were calculated more than once, with the only difference being the storage of the computations in an array. The decoded $\hat{u}_i$ for BSC is estimated as

$$\hat{u}_i \begin{cases} u_i & \text{, if } u_i \text{ is a frozen bit,} \\ 0 & \text{, if } L_N^{(i)}(y_1^N, \hat{u}_1^{i-1}) \geq 1, \\ 1 & \text{, otherwise.} \end{cases}$$

The likelihood ratio is defined as

$$L_N^{(i)}(y_1^N, u_1^{i-1}) = \frac{W_N^{(i)}(y_1^N, u_1^{i-1}|0)}{W_N^{(i)}(y_1^N, u_1^{i-1}|1)}.$$

To estimate $L_N^{(i)}(y_1^N, \hat{u}_1^{i-1})$, we use the recursive formulas

$$L_N^{(2i-1)}(y_1^N, u_1^{2i-2}) = \frac{L_{N/2}^{(i)}(y_1^{N/2}, u_{1,o}^{2i-2} \oplus u_{1,e}^{2i-2}) L_{N/2}^{(i)}(y_{N/2+1}^N, u_{1,e}^{2i-2}) + 1}{L_{N/2}^{(i)}(y_1^{N/2}, u_{1,o}^{2i-2} \oplus u_{1,e}^{2i-2}) + L_{N/2}^{(i)}(y_{N/2+1}^N, u_{1,e}^{2i-2})} \tag{2.17}$$

and

$$L_N^{(2i)}(y_1^N, u_1^{2i-1}) = [L_{N/2}^{(i)}(y_1^{N/2}, u_{1,o}^{2i-2} \oplus u_{1,e}^{2i-2})]^{1-2u_{2i-1}} \cdot L_{N/2}^{(i)}(y_{N/2+1}^N, u_{1,e}^{2i-2}). \tag{2.18}$$

At the Nth level, each calculation of a LR at length N is reduced to the calculation of two LRs at length $N/2$, each calculation of a LR at length $N/2$ is reduced to the calculation of two LRs at length $N/4$ and so on. This recursion can be computed down to block length 1, at which point the LRs have the form

$$L_1^{(1)}(y_1) = \frac{W(y_i|0)}{W(y_i|1)}.$$

2.4 Performance on BEC

In Fig. 2.4, we observe the improvement of the BER performance as the block length grows larger. This happens because perfect channel polarization is accomplished when the block length increases to infinity.

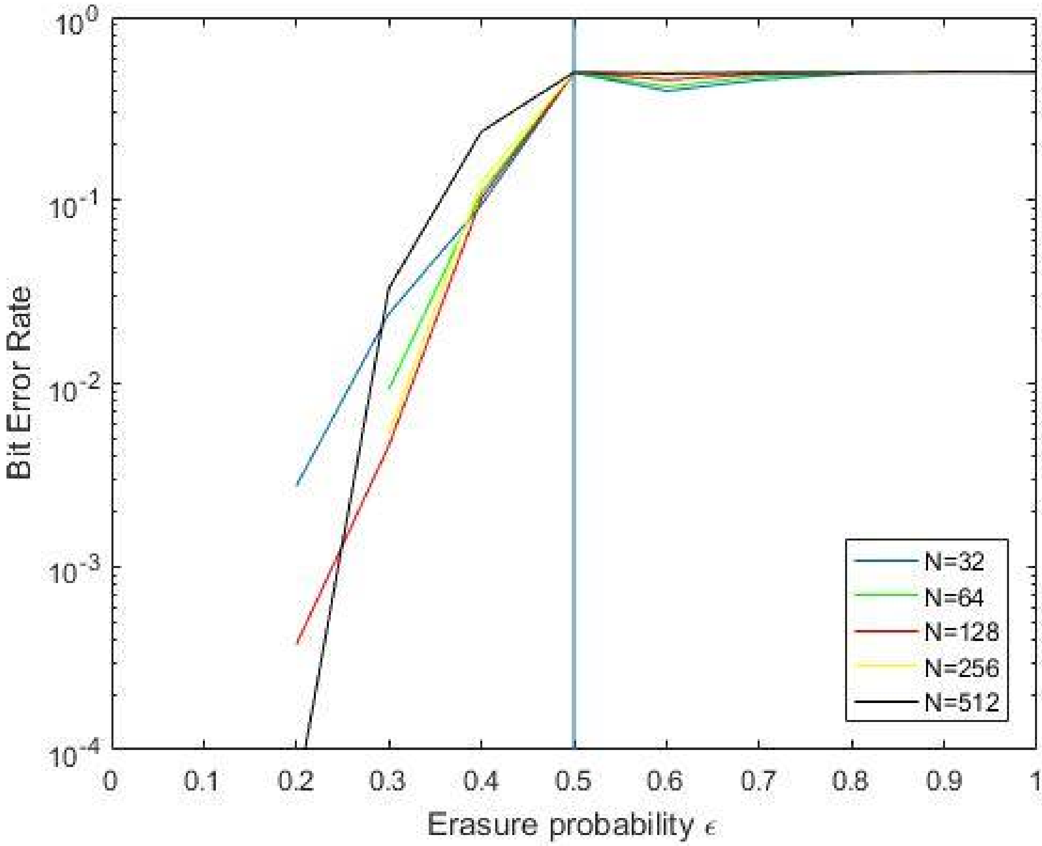


Figure 2.9: BER of transmissions over BEC, with rate = 0.5.

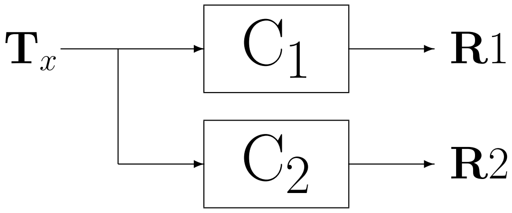


Figure 2.10: Signal model

2.5 Application of Polar Coding to the Degraded Broadcast Binary Erasure Channel

In Fig. 2.10, we have a transmitter Tx and two receivers Rx_1 and Rx_2. Between the transmitter and the receivers there are two communication channels C_1 and C_2. They are both BECs with erasure probability $\epsilon_1 = 0.25$ and and $\epsilon_2 = 0.5$ and capacity equal to 0.75 and 0.5 accordingly.

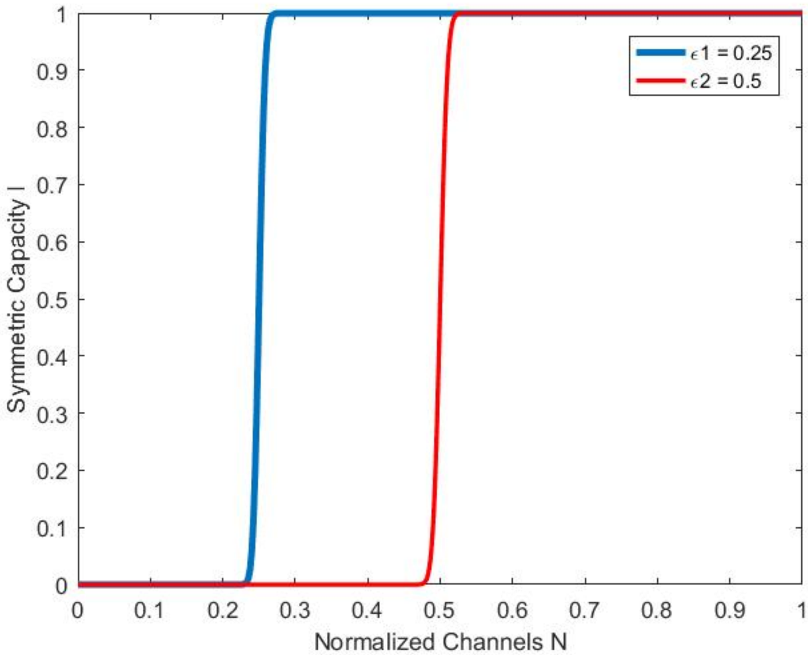


Figure 2.11: Symmetric Capacities of new virtual channels $N = 2^{20}$.

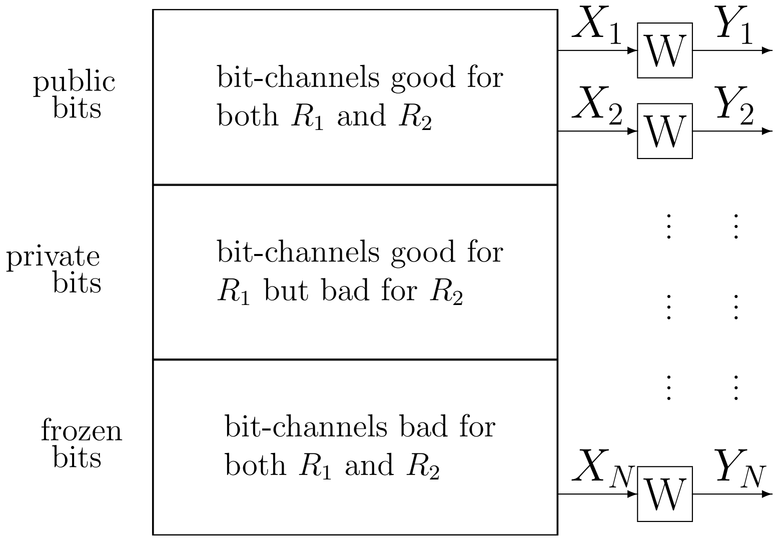


Figure 2.12: Encoding.

After polar encoding of $N = 2^{20}$ at the transmitter, 2^{20} new virtual channels between the transmitter and each receiver are created. We can recursively calculate the symmetric capacities of the new virtual channels of Rx_1 and Rx_2, using the formulas (2.12), (2.13) with $I(W) = 1-\epsilon$. Then we sort them and plot them in Fig. 2.11, where we observe that the 75% of the virtual channels of the first receiver have capacity equal to 1. The capacities of the second receiver's new virtual channels are 50% equal to 1 and the other 50% equal to 0. The transmitter takes advantage of the fact that, between 0.25 and 0.5, only the capacities of the first receiver are equal to 1 and sends the private data with rate $R_{private} = C_1 - C_2$. This way, only the first receiver can decode successfully the private information, while the capacities of the second one are 0. When the transmitter needs to send public data to both, it uses rate $R_{public} = C_2$, so that both Rx_1 and Rx_2 can recover it.

As shown in Fig. 2.12, we used the encoding of polar codes to transmit the private data only over those bit-channels $W_N^{(i)}$ that are bad for Rx_2, public information to the bit-channels that are good for both Rx_1 and Rx_2, and frozen bits to those that are bad for both Rx_1 and Rx_2.

Appendix

Transition Probabilities of BSC for $I(u_1; y_1, y_2)$:

- $p(y_1 = 0, y_2 = 0|u_1 = 0) = \frac{1}{2}(1-p)^2 + \frac{1}{2}p^2$
- $p(y_1 = 0, y_2 = 0|u_1 = 1) = \frac{1}{2}(1-p)p + \frac{1}{2}p(1-p)$
- $p(y_1 = 0, y_2 = 1|u_1 = 0) = \frac{1}{2}(1-p)p + \frac{1}{2}p(1-p)$
- $p(y_1 = 0, y_2 = 1|u_1 = 1) = \frac{1}{2}(1-p)^2 + \frac{1}{2}p^2$
- $p(y_1 = 1, y_2 = 0|u_1 = 0) = \frac{1}{2}(1-p)p + \frac{1}{2}p(1-p)$
- $p(y_1 = 1, y_2 = 0|u_1 = 1) = \frac{1}{2}(1-p)^2 + \frac{1}{2}p^2$
- $p(y_1 = 1, y_2 = 1|u_1 = 0) = \frac{1}{2}(1-p)^2 + \frac{1}{2}p^2$
- $p(y_1 = 1, y_2 = 1|u_1 = 1) = \frac{1}{2}(1-p)p + \frac{1}{2}p(1-p)$

Transition Probabilities of BSC for $I(u_2; y_1, y_2, u_1)$:

- $p(y_1 = 0, y_2 = 0, u_1 = 0|u_2 = 0) = \frac{1}{2}(1-p)^2$
- $p(y_1 = 0, y_2 = 0, u_1 = 0|u_2 = 1) = \frac{1}{2}p^2$
- $p(y_1 = 0, y_2 = 0, u_1 = 1|u_2 = 0) = \frac{1}{2}(1-p)p$
- $p(y_1 = 0, y_2 = 0, u_1 = 1|u_2 = 1) = \frac{1}{2}(1-p)p$
- $p(y_1 = 0, y_2 = 1, u_1 = 0|u_2 = 0) = \frac{1}{2}(1-p)p$
- $p(y_1 = 0, y_2 = 1, u_1 = 0|u_2 = 1) = \frac{1}{2}(1-p)p$
- $p(y_1 = 0, y_2 = 1, u_1 = 1|u_2 = 0) = \frac{1}{2}p^2$
- $p(y_1 = 0, y_2 = 1, u_1 = 1|u_2 = 1) = \frac{1}{2}(1-p)^2$
- $p(y_1 = 1, y_2 = 0, u_1 = 0|u_2 = 0) = \frac{1}{2}(1-p)p$
- $p(y_1 = 1, y_2 = 0, u_1 = 0|u_2 = 1) = \frac{1}{2}(1-p)p$
- $p(y_1 = 1, y_2 = 0, u_1 = 1|u_2 = 0) = \frac{1}{2}(1-p)^2$

- $p(y_1 = 1, y_2 = 0, u_1 = 1|u_2 = 1) = \frac{1}{2}p^2$
- $p(y_1 = 1, y_2 = 1, u_1 = 0|u_2 = 0) = \frac{1}{2}p^2$
- $p(y_1 = 1, y_2 = 1, u_1 = 0|u_2 = 1) = \frac{1}{2}(1-p)^2$
- $p(y_1 = 1, y_2 = 1, u_1 = 1|u_2 = 0) = \frac{1}{2}(1-p)p$
- $p(y_1 = 1, y_2 = 1, u_1 = 1|u_2 = 1) = \frac{1}{2}(1-p)p$

Transition Probabilities of BEC for $I(u_1; y_1, y_2)$:

- $p(y_1 = 0, y_2 = 0|u_1 = 0) = \frac{1}{2}(1-\epsilon)^2$
- $p(y_1 = 0, y_2 = 0|u_1 = 1) = 0$
- $p(y_1 = 0, y_2 = e|u_1 = 0) = \frac{1}{2}(1-\epsilon)\epsilon$
- $p(y_1 = 0, y_2 = e|u_1 = 1) = \frac{1}{2}(1-\epsilon)\epsilon$
- $p(y_1 = 0, y_2 = 1|u_1 = 0) = 0$
- $p(y_1 = 0, y_2 = 1|u_1 = 1) = \frac{1}{2}(1-\epsilon)^2$
- $p(y_1 = e, y_2 = 0|u_1 = 0) = \frac{1}{2}(1-\epsilon)\epsilon$
- $p(y_1 = e, y_2 = 0|u_1 = 1) = \frac{1}{2}\epsilon^2 + \frac{1}{2}\epsilon^2$
- $p(y_1 = e, y_2 = e|u_1 = 0) = \frac{1}{2}\epsilon^2 + \frac{1}{2}\epsilon^2$
- $p(y_1 = e, y_2 = e|u_1 = 1) = \frac{1}{2}(1-\epsilon)\epsilon$
- $p(y_1 = e, y_2 = 1|u_1 = 0) = \frac{1}{2}(1-\epsilon)\epsilon$
- $p(y_1 = e, y_2 = 1|u_1 = 1) = 0$
- $p(y_1 = 1, y_2 = 0|u_1 = 0) = 0$
- $p(y_1 = 1, y_2 = 0|u_1 = 1) = \frac{1}{2}(1-\epsilon)^2$
- $p(y_1 = 1, y_2 = e|u_1 = 0) = \frac{1}{2}(1-\epsilon)\epsilon$
- $p(y_1 = 1, y_2 = e|u_1 = 1) = \frac{1}{2}(1-\epsilon)\epsilon$
- $p(y_1 = 1, y_2 = 1|u_1 = 0) = \frac{1}{2}(1-\epsilon)^2$
- $p(y_1 = 1, y_2 = 1|u_1 = 1) = 0$

Transition Probabilities of BEC for $I(u_2; y_1, y_2, u_1)$:

- $p(y_1 = 0, y_2 = 0, u_1 = 0|u_2 = 0) = \frac{1}{2}(1-\epsilon)^2$
- $p(y_1 = 0, y_2 = 0, u_1 = 0|u_2 = 1) = 0$
- $p(y_1 = 0, y_2 = 0, u_1 = 1|u_2 = 0) = 0$
- $p(y_1 = 0, y_2 = 0, u_1 = 1|u_2 = 1) = 0$
- $p(y_1 = 0, y_2 = e, u_1 = 0|u_2 = 0) = \frac{1}{2}(1-\epsilon)\epsilon$
- $p(y_1 = 0, y_2 = e, u_1 = 0|u_2 = 1) = 0$
- $p(y_1 = 0, y_2 = e, u_1 = 1|u_2 = 0) = 0$
- $p(y_1 = 0, y_2 = e, u_1 = 1|u_2 = 1) = \frac{1}{2}(1-\epsilon)\epsilon$
- $p(y_1 = 0, y_2 = 1, u_1 = 0|u_2 = 0) = 0$
- $p(y_1 = 0, y_2 = 1, u_1 = 0|u_2 = 1) = 0$
- $p(y_1 = 0, y_2 = 1, u_1 = 1|u_2 = 0) = 0$
- $p(y_1 = 0, y_2 = 1, u_1 = 1|u_2 = 1) = \frac{1}{2}(1-\epsilon)^2$
- $p(y_1 = e, y_2 = 0, u_1 = 0|u_2 = 0) = \frac{1}{2}(1-\epsilon)\epsilon$
- $p(y_1 = e, y_2 = 0, u_1 = 0|u_2 = 1) = 0$
- $p(y_1 = e, y_2 = 0, u_1 = 1|u_2 = 0) = \frac{1}{2}(1-\epsilon)\epsilon$
- $p(y_1 = e, y_2 = 0, u_1 = 1|u_2 = 1) = 0$
- $p(y_1 = e, y_2 = e, u_1 = 0|u_2 = 0) = \frac{1}{2}\epsilon^2$
- $p(y_1 = e, y_2 = e, u_1 = 0|u_2 = 1) = \frac{1}{2}\epsilon^2$
- $p(y_1 = e, y_2 = e, u_1 = 1|u_2 = 0) = \frac{1}{2}\epsilon^2$
- $p(y_1 = e, y_2 = e, u_1 = 1|u_2 = 1) = \frac{1}{2}\epsilon^2$
- $p(y_1 = e, y_2 = 1, u_1 = 0|u_2 = 0) = 0$
- $p(y_1 = e, y_2 = 1, u_1 = 0|u_2 = 1) = \frac{1}{2}(1-\epsilon)\epsilon$
- $p(y_1 = e, y_2 = 1, u_1 = 1|u_2 = 0) = 0$
- $p(y_1 = e, y_2 = 1, u_1 = 1|u_2 = 1) = \frac{1}{2}(1-\epsilon)\epsilon$

- $p(y_1 = 1, y_2 = 0, u_1 = 0|u_2 = 0) = 0$
- $p(y_1 = 1, y_2 = 0, u_1 = 0|u_2 = 1) = 0$
- $p(y_1 = 1, y_2 = 0, u_1 = 1|u_2 = 0) = \frac{1}{2}(1 - \epsilon)^2$
- $p(y_1 = 1, y_2 = 0, u_1 = 1|u_2 = 1) = 0$
- $p(y_1 = 1, y_2 = e, u_1 = 0|u_2 = 0) = 0$
- $p(y_1 = 1, y_2 = e, u_1 = 0|u_2 = 1) = \frac{1}{2}(1 - \epsilon)\epsilon$
- $p(y_1 = 1, y_2 = e, u_1 = 1|u_2 = 0) = \frac{1}{2}(1 - \epsilon)\epsilon$
- $p(y_1 = 1, y_2 = e, u_1 = 1|u_2 = 1) = 0$
- $p(y_1 = 1, y_2 = 1, u_1 = 0|u_2 = 0) = 0$
- $p(y_1 = 1, y_2 = 1, u_1 = 0|u_2 = 1) = \frac{1}{2}(1 - \epsilon)^2$
- $p(y_1 = 1, y_2 = 1, u_1 = 1|u_2 = 0) = 0$
- $p(y_1 = 1, y_2 = 1, u_1 = 1|u_2 = 1) = 0$

Proof of $I(W_4{}^{(4)}) = 2I(W_2{}^{(2)}) - I(W_2{}^{(2)})^2$*:*

$$
\begin{aligned}
Z(W_4{}^{(4)}) &= \sum_{y_1^4, u_1^3} \sqrt{W_4{}^{(4)}(y_1^4, u_1^3|0)}\sqrt{W_4{}^{(4)}(y_1^4, u_1^3|1)} \\
&= \sum_{y_1^4, u_1^3} \frac{1}{8}\sqrt{W(y_1|u_1 \oplus u_2 \oplus u_3 \oplus 0)W(y_2|u_3 \oplus 0)W(y_3|u_2 \oplus 0)W(y_4|0)} \\
&\cdot \sqrt{W(y_1|u_1 \oplus u_2 \oplus u_3 \oplus 1)W(y_2|u_3 \oplus 1)W(y_3|u_2 \oplus 1)W(y_4|1)} \\
&= \frac{1}{8}\sum_{y_1, u_1^3}(\sqrt{W(y_1|u_1 \oplus u_2 \oplus u_3)W(y_1|u_1 \oplus u_2 \oplus u_3 \oplus 1)}) \cdot \sum_{y_2} \sqrt{W(y_2|u_3)W(y_2|u_3 \oplus 1)} \\
&\cdot \sum_{y_3} \sqrt{W(y_3|u_2)W(y_3|u_2 \oplus 1)} \cdot \sum_{y_4} \sqrt{W(y_4|0)W(y_4|1)} \\
&= \frac{1}{8}(\sum_{y_1, u_1^2} \sqrt{W(y_1|u_1 \oplus u_2 \oplus 0)W(y_1|u_1 \oplus u_2 \oplus 1)} + \\
&+ \sum_{y_1, u_1^2} \sqrt{W(y_1|u_1 \oplus u_2 \oplus 1)W(y_1|u_1 \oplus u_2 \oplus 0)}) \cdot Z(W) \cdot Z(W) \cdot Z(W) \\
&= \frac{1}{8}(\sum_{y_1^2, u_1} \sqrt{W(y_1|u_1 \oplus 0)W(y_1|u_1 \oplus 1)} + \sum_{y_1, u_1} \sqrt{W(y_1|u_1 \oplus 1)W(y_1|u_1 \oplus 0)} + \\
&+ \sum_{y_1, u_1} \sqrt{W(y_1|u_1 \oplus 0)W(y_1|u_1 \oplus 1)} + \sum_{y_1, u_1} \sqrt{W(y_1|u_1 \oplus 1)W(y_1|u_1 \oplus 0)}) \cdot Z(W)^3 \\
&= \frac{1}{8} \cdot 4 \sum_{y_1, u_1} \sqrt{W(y_1|u_1)W(y_1|u_1 \oplus 1)} \cdot Z(W)^3 \\
&= \frac{1}{8} \cdot 4(\sum_{y_1} \sqrt{W(y_1|0)W(y_1|0 \oplus 1)} + \sum_{y_1} \sqrt{W(y_1|1)W(y_1|1 \oplus 1)}) \cdot Z(W)^3 \\
&= \frac{1}{8} \cdot 4 \cdot 2Z(W) \cdot Z(W)^3 \\
&= Z(W)^4 \\
&= Z(W_2{}^{(2)})^2
\end{aligned}
$$

$$
\begin{aligned}
Z(W_4{}^{(4)}) = Z(W_2{}^{(2)})^2 &\Leftrightarrow \\
&\Leftrightarrow 1 - I(W_4{}^{(4)}) = (1 - I(W_2{}^{(2)}))^2 \\
&\Leftrightarrow 1 - I(W_4{}^{(4)}) = (1 + I(W_2{}^{(2)})^2 - 2I(W_2{}^{(2)}) \\
&\Leftrightarrow \boldsymbol{I(W_4{}^{(4)}) = 2I(W_2{}^{(2)}) - I(W_2{}^{(2)})^2}
\end{aligned}
$$

Bibliography

[1] E. Arikan. Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. *IEEE Transactions on Information Theory*, 55(7):3051–3073, 2009.

[2] A. Carlton. Surprise! polar codes are coming in from the cold, 2016 (accessed May 2018). `https://www.computerworld.com/article/3151866/mobile-wireless/surprise-polar-codes-are-coming-in-from-the-cold.html`

[3] A. Carlton. How polar codes work, 2017 (accessed May 2018). `https://www.computerworld.com/article/3228804/mobile-wireless/how-polar-codes-work.html#tk.drr_mlt`

[4] R. A. Chou and M. R. Bloch. Polar coding for the broadcast channel with confidential messages: A random binning analogy. *IEEE Transactions on Information Theory*, 62(5):2410–2429, 2016.

[5] Y. Fountzoulas, A. Kosta, and G. N. Karystinos. Polar-code-based security on the bsc-modeled harq in fading. In *Telecommunications (ICT), 2016 23rd International Conference on*, 1–5, IEEE, 2016.

[6] C. Leroux, A. J. Raymond, G. Sarkis, I. Tal, A. Vardy, and W. J. Gross. Hardware implementation of successive-cancellation decoders for polar codes. *Journal of Signal Processing Systems*, 69(3):305–315, 2012.

[7] H. Mahdavifar and A. Vardy. Achieving the secrecy capacity of wiretap channels using polar codes. *IEEE Transactions on Information Theory*, 57(10):6428–6443, 2011.

[8] C. E. Shannon. Communication theory of secrecy systems. *Bell Labs Technical Journal*, 28(4):656–715, 1949.

[9] A. D. Wyner. The wire-tap channel. *Bell Labs Technical Journal*, 54(8):1355–1387, 1975.