KING ABDULLAH UNIVERSITY OF SCIENCE AND TECHNOLOGY

EE-242 DIGITAL COMMUNICATION

TERM PAPER PROJECT

Channel Polarization And Polar Codes

By

Anlei Rao
Fakhruddin Mahmood

For

Dr. Tareq Naffouri

TABLE OF CONTENTS

Abstract

Polar Codes, a coding scheme introduced by Arikan, are the first family of codes known to achieve the symmetric capacity of any given binary-input discrete memoryless channel (B-DMC) W by using a low-complexity successive cancellation (SC) decoder. By recursively combining and splitting the independent B-DMCs, some of these polarized channels become error-free or noiseless to achieve the symmetric capacity $\mathrm{I(W)}$, while others are completely noise approach $1-\mathrm{I(W)}$. The polarized channels are well-conditioned for channel coding such that, we just need to send the information bits through these noiseless channels at rate 1 and fixed symbols through the noisy channels at rate 0. Based on the phenomenon of channel polarization, we can construct the polar codes with the block-length N in powers of 2 at a rate higher than the target rate $\mathrm{R} \leq \mathrm{I(W)}$. The probability of block error will be bounded by $\mathrm{P_e} = \mathrm{O}(\mathrm{N}^{-\frac{1}{4}})$ independent of the code rate with the coding and decoding complexity $\mathrm{O(NlogN)}$ for each.

I. Introduction

Polar codes, a coding scheme that was first introduced by Arikan in [1], can achieve the Shannon capacity of arbitrary symmetric B-DMC W under low encoding and decoding complexity. The main idea of polar codes is based on the phenomenon of "channel polarization". More precisely, by recursively combining and splitting individual channels, some of these channels become essentially error-free, while others become completely noise. Further, the fraction of the noiseless channels tends to the capacity of the underlying binary symmetric channels.

The element for channel polarization is a B-DMC $\mathrm{W}: \mathrm{X} \rightarrow \mathrm{Y}$ with input $\mathrm{X} = \{0, 1\}$, output Y and the transition probabilities P(y|x). Now we can define the mutual information between the input X and output Y as:

$$\mathrm{I}(\mathrm{X}; \mathrm{Y}) \triangleq \sum_{y \in Y} \sum_{x \in X} \mathrm{P}(x)\mathrm{P}(y|x) \log \frac{\mathrm{P}(y|x)}{\sum_{x \in X} \mathrm{P}(x)\mathrm{W}(y|x)} \qquad (1)$$

The Shannon capacity gives the maximum value of mutual information given any input, so: $\mathrm{C} = \max\{\mathrm{I}(\mathrm{X}; \mathrm{Y})\}$.

On the other hand, when X={0, 1} are transmitted with equal probability, which means P(0)=P(1)=$\frac{1}{2}$, we get the symmetric capacity of the channel as:

$$\mathrm{I}(\mathrm{W}) \triangleq \sum_{y \in Y} \sum_{x \in X} \frac{1}{2} \mathrm{P}(y|x) \log \frac{\mathrm{P}(y|x)}{\sum_{x \in X} \frac{1}{2} \mathrm{P}(y|x)} \qquad (2)$$

The symmetric capacity I(W) is equal to Shannon capacity C under the condition that the B-DMC is symmetric, i.e. the rows of the channel matrix are permutation of a probability set and the columns are the same, such as:

$$P(Y|X) = \begin{bmatrix} 1/3 & 1/3 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/3 & 1/3 \end{bmatrix}$$

The rows of the channel matrix are permutation of {1/3, 1/3, 1/6, 1/6}, and the columns are permutation of {1/3, 1/6}.

Channel polarization synthesizes a set of N binary-input channels $\{W_N^i\}$ out of N independent copies of the symmetric B-DMC W with the block length N in powers of 2. As N grows large, the channels seen by individual bits start polarizing, i.e. they either approach a noiseless channel or a pure-noise channel. The channel polarization phenomenon suggests us to use the noiseless channels for transmitting information while fixing the symbols transmitted through the noisy ones to a value known both to sender as well as receiver.

Polar codes are constructed over the binary field GF(2) with the generator matrix G_N directly obtained from the process of channel combining. With the input bits x_1^N, we get the codeword $u_1^N = x_1^N G_N$ as a linear transformation. At the receiver, the SC decoder will do the decoding work with the block error probability bounded by $O(N^{-\frac{1}{4}})$, and a complexity of $O(N\log N)$ is achievable both for encoding and decoding with sufficiently large N [1].

This paper is organized as follows: Section 2 will focus on channel polarization, which includes channel combining and channel splitting. Based on channel polarization, section 3 will discuss on the encoding and decoding of polar codes, and then briefly analyses the performance of the polar codes under successive cancellation (SC) decoding. And finally Section 4 will draw some significant conclusions and discussions.

II. Channel Polarization

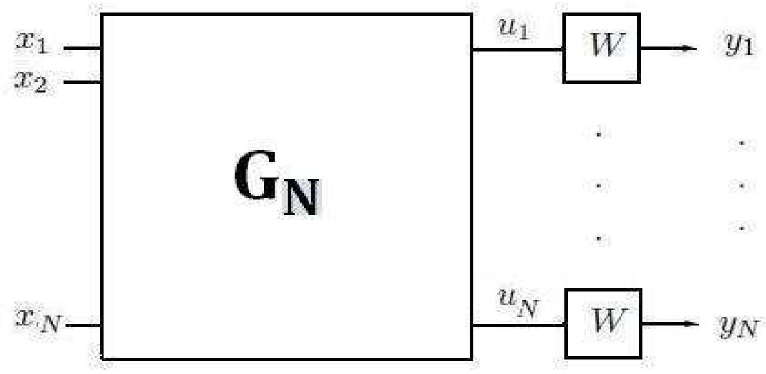
Channel Polarization is defined as an operation that converts N independent copies of a given B-DMC W into another set of N channels $\{W_N^i : 1 \leq i \leq N\}$, called polarized channel set. The channel set $\{W_N^i\}$ polarize in the sense that, as N goes to infinity in powers of 2, the channel capacity $I(W_N^i)$ approach $I(W)$ for the indices $i \in S$, and $1 - I(W)$ for the indices $i \notin S$, where $S \subseteq \{1, 2, 3 \cdots\cdots N\}$ is the information set. Generally, the process of channel polarization can be divided into two parts namely channel combining and channel splitting.

A. Channel Combining

In channel combining, the N independent copies of B-DMC W are combined into an interim channel $W_N: X^N \rightarrow Y^N$ in a recursive manner with the transition probability:

$$P_N(y_1^N|x_1^N) = P^N(y_1^N|x_1^N G_N) \qquad (3)$$

where G_N is the generator matrix, and the general form [2] is:

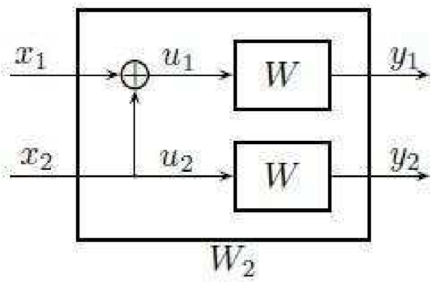


$$W_N: X^N \rightarrow Y^N \quad P_N(y_1^N|x_1^N) = P^N(y_1^N|x_1^N G_N)$$

Fig. 1 The general form of W_N

When $n = 0$ then $N = 2^n = 1$, we define $W_1 = W$.

When $n = 1$ then $N = 2^n = 2$, we construct $W_2: X^2 \rightarrow Y^2$ in the following way:



$$[u_1 \quad u_2] = [x_1 \quad x_2]\begin{bmatrix}1 & 0\\ 1 & 1\end{bmatrix}$$

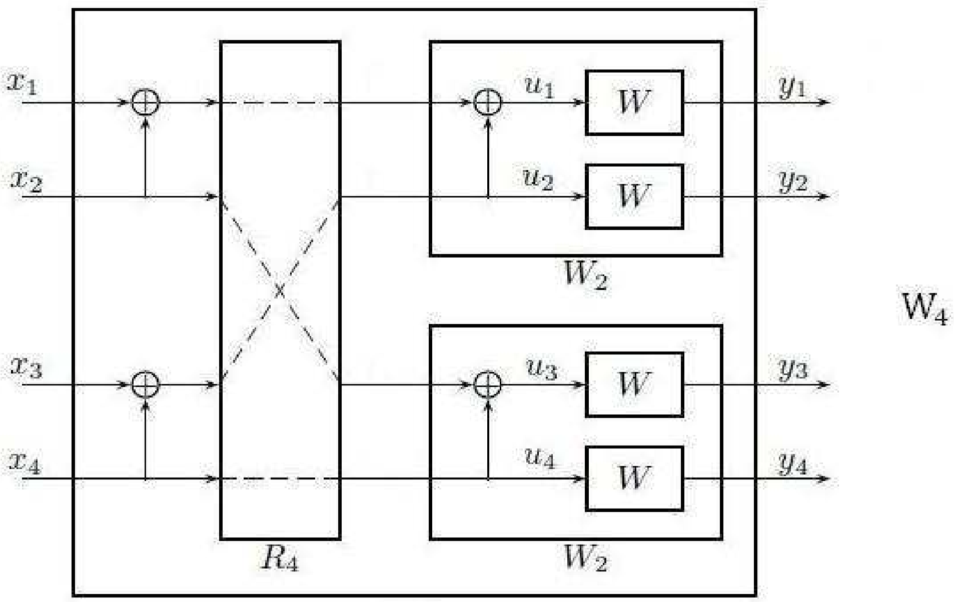
Fig. 2 The Channel W_2

From the construction we can get:

$$P_2(y_1^2|x_1^2) = P(y_1|x_1 \oplus x_2)P(y_2|x_2) = P^2(y_1^2|x_1^2 G_2)$$

With the generator matrix given by $G_2 = \begin{bmatrix}1 & 0\\ 1 & 1\end{bmatrix}$.

When $n = 2$ then $N = 2^n = 4$, we construct $W_4: X^4 \to Y^4$ in a recursive way from W_2:



$$[u_1 \quad u_2 \quad u_3 \quad u_4] = [x_1 \quad x_2 \quad x_3 \quad x_4]\begin{bmatrix}1 & 0 & 0 & 0\\ 1 & 0 & 1 & 0\\ 1 & 1 & 0 & 0\\ 1 & 1 & 1 & 0\end{bmatrix}$$

Fig. 3 The Channel W_4 based on W_2

In the above figure, R_4 is the permutation operation that maps the indices of one set $\{1, 2, 3, 4\}$ to another set $\{1, 3, 2, 4\}$.

In the same way from the construction, we can get:

$$P_4(y_1^4|x_1^4) = P_2(y_1^2|x_1 \oplus x_2, x_3 \oplus x_4)P_2(y_3^4|x_2, x_4) = P^4(y_1^4|x_1^4 G_4)$$

With the generator matrix given by $G_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix}$.

Then generally for any n>0, W_N can be constructed from $W_{N/2}$ in the same way of constructing W_4 from W_2:

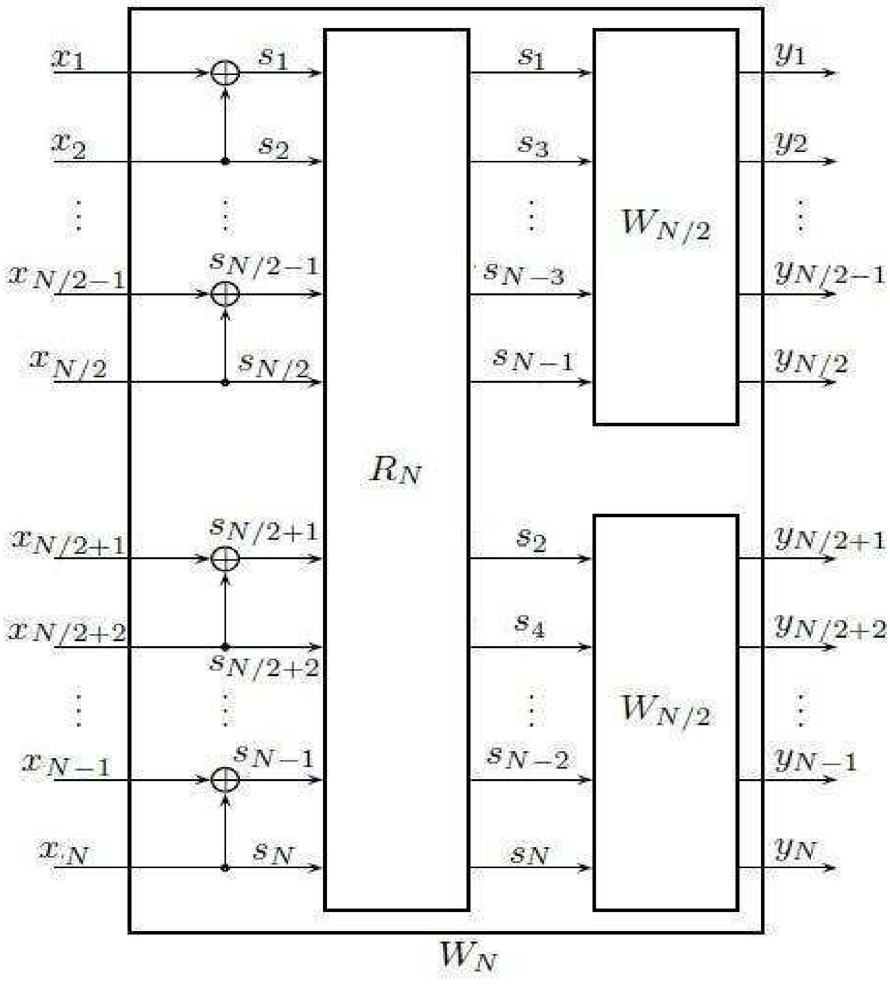


Fig. 4 Recursive construction of W_N from two copies of $W_{N/2}$

The permutation operation R_N just maps the indices of the set $\{1, 2,3 \cdots\cdots N\}$ to $\{1, 3 \cdots N-1, 2, 4 \cdots N\}$.

From the recursive process of constructing W_N from $W_{N/2}$, we can get the relationship of the generator matrix G_N and $G_{N/2}$ as:

$$G_N = \left(I_{\frac{N}{2}} \otimes G\right) R_N \left(I_2 \otimes G_{\frac{N}{2}}\right) \qquad (4)$$

where I_k is the k×k identity matrix and $\otimes$ is the Kronecker product defined as follows:

Kronecker product of two matrix $A = [a_{ij}]_{m\times n}$ and $B_{r\times s}$ is:

$$A\otimes B = \begin{bmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{n1}B & \dots & a_{nn}B \end{bmatrix}_{mr\times ns}$$

By some algebraic calculations, we get:

$$(I_{N/2} \otimes G)R_N = R_N(I_2 \otimes G_{N/2})$$

then we can write G_N in a simpler way:

$$G_N = R_N \left(I_{\frac{N}{2}} \otimes G\right) \left(I_2 \otimes G_{\frac{N}{2}}\right) \qquad (5)$$

With this form, we can get an alternative realization of the recursive construction of G_N:

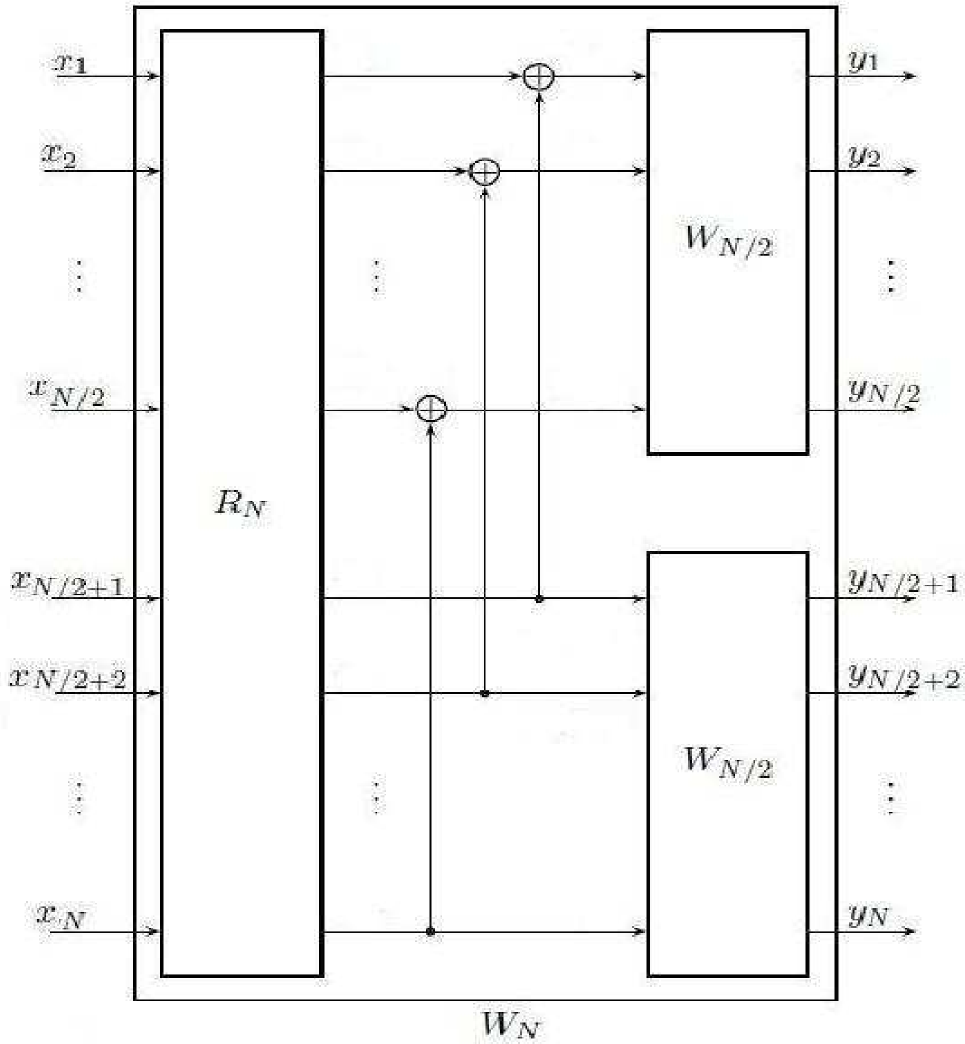


Fig. 5 An alternative realization of recursive construction for W_N

Actually, the transform $I_{N/2} \otimes G$ and R_N are linear, so by just changing the order of them, we can easily get the Fig.4 of W_N from Fig.3.

B. Channel splitting

In the first phase, we combined N independent copies of B-DMC W into the channel $W_N: X^N \rightarrow Y^N$. In this phase we will split $W_N: X^N \rightarrow Y^N$ into the polarized channel set $\{W_N^i: 1 \leq i \leq N\}$ with the transition probabilities defined as follows:

$$P_N^i\left(y_1^N, x_1^{i-1} \middle| x_i\right) \triangleq \sum_{x_1^{I-1}} \frac{1}{2^{N-1}} P_N\left(y_1^N \middle| x_1^N\right) \qquad (6)$$

After splitting, the single channel W_N^i in the set has input x^i and output (y_1^N, x_1^{i-1}) with the form: $W_N^i: X \to Y^N \times X^{i-1}$.

First we consider the situation of N=2 with $W_2: X^2 \to Y^2$. The mutual information can be obtained from the definition (1):

$$I(x_1^2;\ y_1^2) = I(u_1^2;\ y_1^2) = 2I(W)$$

Then we split it into two parts:

$$I(x_1^2;\ y_1^2) = I(x_2;\ y_1^2) + I(x_2;\ y_1^2|x_1) = I(x_1;\ y_1^2) + I(x_2;\ y_1^2, x_1)$$

We interpret the first part $I(x_1;\ y_1^2)$ as the mutual information of the channel between x_1 and y_1^2, with x_2 considered as noise, which we denote as W_2^1. The second part $I(x_2;\ y_1^2, x_1)$ can be denoted as the mutual information of the channel between x_2 and y_1^2, x_1, which we denote as W_2^2 [3]. The transition probabilities of the splitted channels are given as:

$$\begin{cases} P_2^1(y_1^2|x_1) = \sum_{x_1} \frac{1}{2} P_2(y_1^2|x_1^2) = \sum_{x_1} \frac{1}{2} P(y_1|x_1 \oplus x_2) P(y_2|x_2) \\ P_2^2(y_1^2, x_1|x_2) = \frac{1}{2} P_2(y_1^2|x_1^2) = \frac{1}{2} P(y_1|x_1 \oplus x_2) P(y_2|x_2) \end{cases}$$

By this, we split the channel W_2 into the channel set $\{W_2^1, W_2^2\}$. As x_2 is considered as noise, the W_2^1 is set to be the error-free channel while W_2^2 is noisy.

When we consider the situation of $N = 2^n$, we can do the splitting with the probability transition given by (6).

III. Polar Codes

The channel polarization phenomenon suggests that we can use the noiseless channels whose $I(W_N^i)$ is near the symmetric capacity $I(W)$ to transmit the information and fix the symbols transmitted in the noise channels whose $I(W_N^i)$ is near 0.

A. Polar Coding

The polar codes are constructed by the generator matrix:

$$u_1^N = x_1^N G_N$$

By selecting the information set $S \subseteq \{1, 2, 3 \cdots\cdots N\}$, we can write it in the following way:

$$u_1^N = x_S G_N(S) \oplus x_{S^c} G_N(S^c) \qquad (7)$$

Where $G_N(S)$ denotes the submatrix of G_N formed by the rows with the indices in S.

Example, for N=4, $S = \{2,4\}$, $x_{S^c} = (1,0)$, then:

$$u_1^2 = (x_2, x_4)\begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \oplus (1,0)\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

By denoting K=|S|, the polar codes can be identified by (N, K, S, x_{S^c}) with code rate K/N.

Since K and N are fixed, the next step is to determine S and x_{S^c}. Usually, the indices of S are chosen as a K-element subset of {1, 2, 3⋯⋯ N} such that:

$$I\left(W_N^i\right) \geq I\left(W_N^j\right), \text{for all } i \in S \text{ and } j \in S^c$$

In this way, polar codes are channel-specific design such that different channel W will give rise to different polar codes.

Here x_{S^c}, which is transmitted in the pure-noise channel and referred to as frozen bits, can be chosen at will. Actually, any choice for x_{S^c} will give rise to the same performance.

B. Successive Cancellation Decoding

Considering the following channel model:

$$\mathbf{x}_1^N \xrightarrow{\text{coding}} \mathbf{u}_1^N \xrightarrow{\text{channel } W_N} \mathbf{y}_1^N \xrightarrow{\text{decoding}} \hat{\mathbf{x}}_1^N$$

The input bits x_1^N are encoded into the polar codeword u_1^N, then transmitted in the channel W^N. The output y_1^N are first received and then decoded into $\hat{x}_1^N$. The task of decoder is to make an estimation $\hat{x}_1^N$ of x_1^N from y_1^N.

Since the frozen bits x_{S^c} are chosen beforehand and contains no information, we can set $\hat{x}_{S^c} = x_{S^c}$, and use the ML rule to make the decision:

$$\hat{x}_i = \begin{cases} x_i & \text{if } i \in S \\ d_i\left(y_1^N, x_1^{i-1}\right) & \text{if } i \notin S \end{cases} \qquad (8)$$

Where:

$$d_i\left(y_1^N, x_1^{i-1}\right) = \begin{cases} 0 & \text{if } \dfrac{P_N^i\left(y_1^N, x_1^{i-1} \middle| 0\right)}{P_N^i\left(y_1^N, x_1^{i-1} \middle| 1\right)} \geq 1 \\ 1 & \text{if } \dfrac{P_N^i\left(y_1^N, x_1^{i-1} \middle| 0\right)}{P_N^i\left(y_1^N, x_1^{i-1} \middle| 1\right)} \leq 1 \end{cases} \quad (9)$$

This algorithm is called successive cancellation decoding which is based on the channel splitting. Just as its name implies, we need the output y_1^N and the previous estimation value $\hat{x}_1^{i-1}$ to make the current estimation value of $\hat{x}_i$.

We denote P_e as the probability of block error for a (N, K, S, x_{S^c}) polar code. Assuming that each data x_s is sent with the probability of $\frac{1}{2^K}$, the P_e can be calculated as:

$$P_e = \frac{1}{2^K} \sum_{x_s} \sum_{\hat{x}_1^N \neq x_1^N} P_N(y_1^N | x_1^N) \quad (10)$$

It is very difficult to calculate explicitly, but we can get an estimation of it as $P_e = O(N^{-\frac{1}{4}})$.

Another important parameter for performance is the complexity. From the process of encoding and decoding, we can simply figure out that the complexity for encoding and SC decoding are both $O(N\log N)$ as functions of block-length N.

IV. Conclusion and Discussion

By combining and splitting the N independent copies of B-DMCs, we can get a set of polarized channels, some of which are noiseless and others are pure-noise. By sending the information bits only through these noiseless channels while fixing the symbols transmitted through the pure-noise ones, the Shannon capacity of the symmetric B-DMC can be achieved.

Polar codes, based on the phenomenon of channel polarization, are capacity-achieving for any symmetric B-DMC with low encoding and decoding complexity $O(N\log N)$. By selecting the set S of information-bit indices and the frozen bits freely, we can make encoding with the generator matrix G_N, which can be drawn directly from the process of channel combining. But what we must pay attention to is that the polar coding is channel-specific design and different B-DMC W will give rise to different polar codes. Meanwhile, the bit one-by-one decoding algorithm called successive cancellation is drawn from the process of channel splitting. Although the explicit form of block-error probability P_e is difficult to get, it is easy to make an estimation as $P_e = O(N^{-\frac{1}{4}})$.

Usually the block length N is in powers of 2, but when N goes to a power of any other number such as l, then $N = l^n$. In this case we can still use the Arikan's rule to make polar codes with the polarizing matrix G, which is a l×l matrix and has the same form with the G mentioned above. What is more is when the block-length N equals to $\prod_{i=1}^{n} l_i$, then the polar codes can still be constructed from the generating matrix of the form $\otimes_i G_i$, where each G_i is a polarizing matrix of size $l_i \times l_i$ [4]. The encoding and decoding complexities are also given by $O(N\log N)$. Further, although originally introduced for channel coding, polar codes are equally useful for source coding applications, such as lossless and lossy compression. In addition to that, it is also optimal for multi-terminal problems, such as Wyner-Ziv and Gelfand-Pinsker problem [5].

V. Simulation

We did a simulation of generating matrix G_N with Matlab. By inputting the exponent n of the block length $N = 2^n$, we can get the generating matrix G_N by the function of "Generating", which will call the function of "Kronecker" to get the Kronecker product of tow matrix.

Function "Generating":

```
function G=Generating(n)
% This function will generate the generating maatrix with the parameter n
A=[1 0;1 1];
for k=0:1:n-1
    if(k==0)
        G=A;
    else
        B=G;
        G=Kronecker(A,B);
        m=power(2, k+1);
        C=zeros(m,m);
        for i=1:1:m/2
            C(1:m,i)=G(1:m,2*i-1);
        end
        for i=m/2+1:1:m
            C(1:m,i)=G(1:m,2*(i-m/2));
        end
        G=C;
    end
end
```

Function "Kronecker":

```
function K=Kronecker(A,B)

[a, b]=size(A);
[m, n]=size(B);
K=zeros(a*m, b*n);

for i=0:1:a-1
    for j=0:1:b-1
        K(i*m+1:(i+1)*m, j*n+1:(j+1)*n)=A(i+1,j+1).*B;
    end
end
```

References:

[1] E. Arıkan, “Channel polarization: A method for constructing capacityachieving codes for symmetric binary-input memoryless channels,” submitted to IEEE Trans. Inform. Theory, 2008.

[2] S. B. Korada and E. S¸as¸o˘glu, “A class of Transformation that Polarize Symmetric Binary-Input Memoryless Channels”. Available at http://arxiv.org/pdf/0807.3806

[3] S. B. Korada, E. S¸as¸o˘glu, and R. Urbanke, “The Compound Capacity of Polar Codes,” Available at the website of http://arxiv.org/pdf/0937.0486

[4] S. B. Korada, E. S¸as¸o˘glu, and R. Urbanke, “Polar codes: Characterization of exponent, bounds, and constructions,” submitted to IEEE Trans. Inform. Theory, 2009.

[5] S. B. Korada and R. Urbanke, “Polar codes are optimal for lossy source coding,” submitted to IEEE Trans. Inform. Theory, 2009.