Introduction to Information Theory

# Lecture 12 - Polar Codes Part a

*Lecturer: Haim Permuter* *Scribe: Yhonatan Gayer*

## I. Intro to Polar Codes

In this paper, we present an introduction to Polar Codes. Over the course of two lectures, we will explore the effectiveness of this algorithm, which involves manipulating a known channel to enhance capacity and reliability in a non-deterministic manner. Our discussion begins with a motivation for improvement, followed by a step-by-step construction of the initial polarization block using an intuitive example. Subsequently, we generalize this algorithm to encompass multiple blocks and all types of channels. The content of this paper is based on the foundational work by Arikan [1], complemented by the insights from [2] and [3].

## II. Repetition Coding

To improve the accuracy of channel estimation, a proposed approach involves utilizing multiple instances of the same channel. By repeating the use of the channel a total of $n$ times, we can increase our confidence in the correctness of the estimation. This strategy is depicted in Fig. 1 as the Repetition Scheme, where $P$ represents the channel. The scheme can be analyzed in terms of $P_e$ and rate.

To assess the reliability of the channel we examining the error probability $P_e$, which holds $\lim_{n\to\infty} P_e = 0$, This property holds true for most channels, including BSC, BEC, and BCC. regarding the rate, each bit of information is transmitted approximately $n$ times (depending on the specific channel). Consequently, the rate is significantly lower than the capacity, which is given by $C_{total} = nC(P)$.

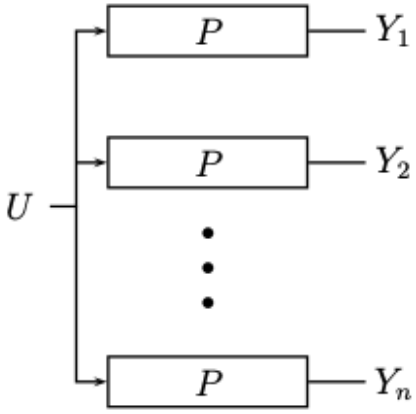


Fig. 1. Repetition Coding Scheme

**Example 1** (Rate and Reliability of Repetition code with $BEC(p)$ Channel) first we will examine the reliability,

$$\begin{aligned}\lim_{n\to\infty} P_e &= \lim_{n\to\infty} P(\hat{U} \neq U) \\ &\overset{(a)}{=} \lim_{n\to\infty} P(Y_1 =?, ..., Y_n =?) \\ &\overset{(b)}{=} \lim_{n\to\infty} (1-p)^n \\ &\overset{(c)}{=} 0\end{aligned} \tag{1}$$

- (a) In BEC channel it's enough that $Y_k \neq ?$ for some $k$ to determine $U$'s value.
- (b) In BEC channel $P(Y_1 =?) = 1 - p$, in a memoryless channel each use of the channel is independant, there for the equality holds.
- (c) The equality hold for $p > 0$.

and now we will examine its rate for a constant $n$,

$$R = \frac{1}{n} << n(1-p) = nC_{BEC} \tag{2}$$

Hence, it can be inferred that Repetition coding offers reliability at the cost of reduced speed.

Now we will show an example of repetition code with $n = 2$, the purpose of this example will be clearer in the next section, where we will use similar scheme with small modifications, and big improvment.

**Example 2** (Repetion Code with $n = 2$, and BEC Channel)

Here is an example of repetition coding with specific parameters: $n = 2$ , $BEC(p)$ channel, and $U \sim Unif(0, 1)$. This scenario is illustrated in Fig.2.

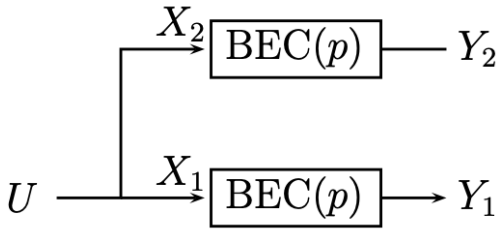


Fig. 2. Repetition code scheme with two BEC channels

Each channel $BEC(p)$ Capcity denotes by $C_{BEC} = 1 - p$ as we saw in the former lecture. The maximum Capcity of two $BEC(p)$ channels denotes by $C_{total} = 2(1 - p)$. The channel's Mutual Information can lower bounded,

$$\begin{aligned} I(U; Y_1, Y_2) &\overset{(a)}{=} I(U; Y_1) + I(U; Y_2|Y_1) \\ &\overset{(b)}{=} C_{BEC} + I(U; Y_2|Y_1) \\ &\geq C_{BEC} \end{aligned} \tag{3}$$

- (a) chain rule
- (b) BEC channel reach capcaity for uniform distrinution as in this case.

and also can upper bounded,

$$\begin{aligned} I(U; Y_1, Y_2) &\overset{(a)}{=} I(X_1, X_2; Y_1, Y_2) \\ &\overset{(b)}{=} I(X_1; Y_1, Y_2) + I(X_2; Y_1, Y_2|X_1) \\ &\leq 2C_{BEC} \end{aligned} \tag{4}$$

- (a) $U = Y_1$ and $U = Y_2$
- (b) chain rule

Overall, from (4) and (3) we can infer that the Described Scheme's Rate does not reach capacity, and has a gap of $2C_{BEC} - I\left(X_2; Y_1, Y_2 | X_1\right)$

## III. Building Block Of Polar Coding

After observing that Repetition coding is not optimal in terms of Rate, we will now propose an alternative solution that incorporates certain modifications, which later on we will see that improves capacity.

**Example 3** (Polar Code with two BEC Channels) We present an example of polar coding with particular parameters: $n = 2$, a BEC with parameter $p$, and two independent uniformly distributed variables $U \sim Unif(0,1)$ and $V \sim Unif(0,1)$. The scenario is depicted in Fig.3, where $X_2 = U \oplus V$ and $X_1 = U$.

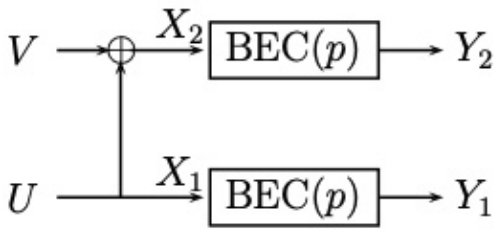


Fig. 3. Polar code scheme with two BEC channels

The channel's Mutual Information can be analyzed in two ways. Firstly, we will examine its value

$$\begin{aligned} I\left(U, V; Y_1, Y_2\right) &\overset{(a)}{=} I\left(X_1, X_2; Y_1, Y_2\right) \\ &\overset{(b)}{=} I\left(X_1; Y_1, Y_2\right) + I\left(X_2; Y_1, Y_2 | X_1\right) \\ &\overset{(c)}{=} I\left(X_1; Y_1, Y_2\right) + I\left(X_2; Y_1, Y_2\right) \\ &\overset{(d)}{=} I\left(X_1; Y_1\right) + I\left(X_2; Y_2\right) \\ &= 2C_{BEC}, \end{aligned} \tag{5}$$

- (a) There is a 1 to 1 transformation from $(U, V)$ to $(X_1, X_2)$.
- (b) Chain rule.
- (c) $X_1$ is independant of $X_2$, since $X_2 = V \oplus U$ and both $U$ and $V$ are $Bernoli(\frac{1}{2})$ and independant.
- (d) $Y_1$ is independant of $X_2$, and $Y_2$ is independant of $X_1$.

which proves that the channel reach it's Capacity.

Now, we will observe the channel's Mutual Information by utilizing the chain rule,

$$I\left(U, V; Y_1, Y_2\right) = I\left(V; Y_1, Y_2\right) + I\left(U; Y_1, Y_2 | V\right) \tag{6}$$

The left term can be further simplified,

$$\begin{aligned} I\left(V; Y_1, Y_2\right) &= H\left(V\right) - H\left(V | Y_1, Y_2\right) \\ &\overset{(a)}{=} 1 - \sum_{y_1=0}^{1} \sum_{y_2=0}^{1} P_{Y_1,Y_2}\left(y_1, y_2\right) H\left(V | Y_1 = y_1, Y_2 = y_2\right) \\ &\overset{(b)}{=} 1 - P_{Y_1,Y_2}\left(y_1 = ?, y_2 = ?\right) H\left(V | Y_1 = ?, Y_2 = ?\right) \\ &\overset{(c)}{=} 1 - p^2 \\ &< C_{BEC} \end{aligned} \tag{7}$$

- (a) $H\left(V\right)$ when $U \sim Unif(0, 1)$.
- (b) When $Y_1 \neq ?$ or $Y_2 \neq ?$ $V$ is known and therefore $H\left(V | Y_1 = y_1, Y_2 = y_2\right) = 0$.
- (c) When $Y_1 = ?$ or $Y_2 = ?$ we have no information on $V$ and therefore $H\left(V | Y_1 = y_1, Y_2 = y_2\right) = 1$ as $V \sim Unif(0, 1)$.

Likewise, the right term can be further reduced,

$$\begin{aligned} I\left(U;Y_1,Y_2|V\right) &\overset{(a)}{=} H\left(U|V\right) - H\left(U|Y_1,Y_2,V\right) \\ &\overset{(b)}{=} H\left(U\right) - \sum_{y_1=0}^{1}\sum_{y_2=0}^{1} P_{Y_1,Y_2}\left(y_1,y_2\right) H\left(U|Y_1=y_1,Y_2=y_2,V\right) \\ &\overset{(c)}{=} 1 - P_{Y_1,Y_2}\left(y_1=x_1,y_2=?\right) - P_{Y_1,Y_2}\left(y_1=?,y_2=x_2\right) - P_{Y_1,Y_2}\left(y_1=?,y_2=?\right) \\ &\overset{(d)}{=} 1-(1-p)p-p(1-p)-p^2 \\ &= (1-p)^2 \\ &> C_{BEC} \end{aligned} \tag{8}$$

- (a) Chain rule.
- (b) $V$ is independant with $V$.
- (c) The only case where $U$ is known is when we know both $Y_1$ and $Y_2$, in that case $H\left(U|Y_1=y_1,Y_2=y_2,V\right)=0$, in all other cases $H\left(U|Y_1=y_1,Y_2=y_2,V\right)=1$.
- (d) $P_{Y_1,Y_2}\left(y_1,y_2\right)=P_{Y_1}(y_1)P_{y_2}(y_2)$.

By referring to equation (6), we can interpret the Two-Channel Scheme depicted in Fig.3 as an equivalent scheme composed of two parallel channels (using that $BEC(p)$ channel has a capacity of $C_{BEC}=1-p$):

- The $W^-$ Channel, which corresponds to a BEC with parameter $p^2$.
- The $W^+$ Channel, which corresponds to a BEC with parameter $1-(1-p)^2$.

The Equivalent Scheme described in Fig. 4

The Equivalent Scheme illustrated in Fig. 4 demonstrates the processing of independent variables $U$ and $V$ to generate the output sequences $Y_1$ and $Y_2$.

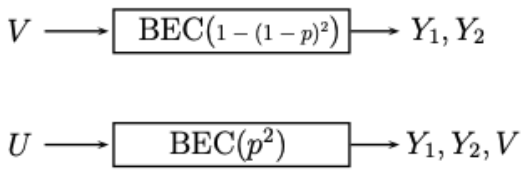


Fig. 4. Equivalent Scheme to the Scheme in Fig.3

The comparison yields four significant conclusions:

- The capacity of Channel $W^-$ is lower than $C_{BEC}$.
- The capacity of Channel $W^+$ is greater than $C_{BEC}$.
- Decoding of Channel $W^-$ involves knowing of $Y_1$ and $Y_2$.
- Decoding of Channel $W^+$ involves knowing of $Y_1$, $Y_2$, and $U$.

Based on our analysis of the scheme depicted in Fig. 3, we conclude that it can be effectively represented as two parallel channels, namely $W^-$ and $W^+$. This finding is noteworthy as it allows us to create a channel with a lower capacity compared to the original $BEC(p)$ channel, while maintaining the same overall scheme capacity.

## IV. Polar Coding of size $n$

In the previous section, we demonstrated the enhancement of capacity through Polar Coding for the case where $n = 2$ and using the $BEC(p)$ channel. In this section, we will generalize this approach to all channels and for all $n = 2^m$ with $m \geq 1$.

By replicating the process outlined in the aforementioned example, we can further enhance the capacities of individual channels while simultaneously diminishing others. This enables us to effectively analyze and utilize channels with higher capacities as data channels, while employing the remaining channels as frozen-bits channels, for instance in the case of $n = 4$ we will get 4 channels, $W^{++}, W^{+-}, W^{-+}$,and $W^{--}$ ,where the signs $+$ and $-$ relates to the capacity of the channels. The channels with enlarged capacities will be counted as data channels, and the channels with low capacities will be counted as frozen-bits channels.

To facilitate understanding, the decoding process for the case of $N = 4$ is thoroughly explained in both Fig. 5 and in (9).

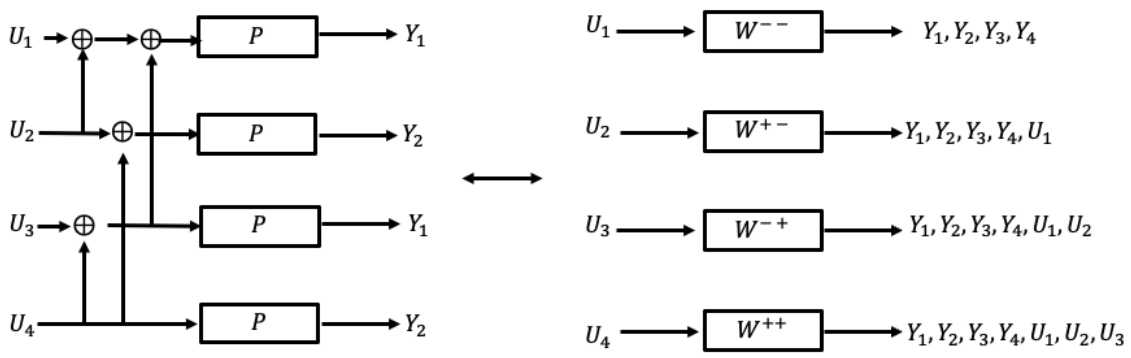


Fig. 5. Polar Coding, N=4

$$
\begin{aligned}
X_1 &= U_1 \oplus U_2 \oplus U_3 \oplus U_4 \\
X_2 &= U_2 \oplus U_4 \\
X_3 &= U_3 \oplus U_4 \\
X_4 &= U_4
\end{aligned}
\tag{9}
$$

**Theorem 1** As the number of the channels grows, the capacity of each channel converges to 1 or 0:

$$
\begin{aligned}
&\forall \delta > 0 : \\
&\lim_{n\to\infty} \frac{\left| C\left(P_i^{(n)}\right) > 1-\delta \right|}{n} = C\left(P\right) \\
&\lim_{n\to\infty} \frac{\left| C\left(P_i^{(n)}\right) < \delta \right|}{n} = 1 - C\left(P\right)
\end{aligned}
\tag{10}
$$

Here, $n$ is the number of channels, $P$ is the original channel, $P_i^{(n)}$ is the $i'th$ channel in the equivalent form, and $C(P)$ represents the capacitty of channel $P$.

The proof of Theorem 1 is out of the scope of this course.

Theorem 1 states that we can approach the capacity $C = 1$ in certain channels as closely as desired, while in other channels we approach the capacity $C = 0$, provided that we choose a sufficiently large value for $N$.

We can utilize the data channels with $C \approx 1$, to send data, while in the frozen-bits channels with $C \approx 0$ we will send $0$ as it's doesn't matter what will be pass through them, as long the decoder knows what are the values of them. It is important to note that these frozen-bits channels must still be utilized, as they play a crucial role in the decoding process of the data channels.

Due to the mathematical complexity associated with calculating the capacity of each channel, particularly for high orders of $n$, where the expressions become exponentially intricate, resorting to Monte Carlo experiments is an acceptable approach to estimate the capacity or reliability of each channel.

## V. Polar Encoding

Polar Encoding focuses on efficiently and mathematically simplifying the arrangement of a bit vector $U = [U_1, ..., U_N]$ into a vector of the same size $N$ $X = [X_1, ..., X_N]$.

In order to formulated Polar encoding mathematically, we will first make some new definitions:

**Definition 1** The operation $\odot$ operates between two matrices as explained here:

$$A \odot B = \begin{bmatrix} [A]_{0,0} B & \dots & [A]_{0,N-1} B \\ \vdots & \ddots & \vdots \\ [A]_{N-1,N-1} B & \dots & [A]_{N-1,N-1} B \end{bmatrix} \tag{11}$$

Where matrix $A$ is of size $N \times N$ and $[A]_{i,j}$ is the $i, j'th$ element of matrix $A$ $\forall 0 \leq i, j \leq N-1$

**Definition 2** Now let's define the binary matrix $G_N$, where $N = 2^m \forall 1 \leq m$. In the case of $N = 2$,

$$G_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}. \tag{12}$$

In the case of $N = 2^m, G_N = G_{N/2} \odot G_{N/2}$ $\forall 2 \leq m$. also note that $G_N$ operates as a XOR between the components as demonstrates in (13):

$$\begin{bmatrix} X_1, X_2 \end{bmatrix} = \begin{bmatrix} U, V \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} U \oplus V, V \end{bmatrix} \tag{13}$$

Using the definitions above, makes it easier to decode $U = [U_1, ..., U_N]$ into $X = [X_1, ..., X_N]$ ,as an example the case of $N = 4$ is presented in (15).

$$X = UG_4 \tag{14}$$

$$\Big[X_1, X_2, X_3, X_4\Big] = \Big[U_1, U_2, U_3, U_4\Big] \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} U_1 \oplus U_2 \oplus U_3 \oplus U_4 \\ U_2 \oplus U_4 \\ U_3 \oplus U_4 \\ U_4 \end{bmatrix}^T \tag{15}$$

## VI. Polar Decoding

Once the data passes through the channel, it becomes necessary to extract the vector of bits denoted as $R = [R_1, ..., R_N]$, which represents the channel outputs. This vector needs to be decoded in the most optimal manner to obtain the estimated vector $\hat{U} = [\hat{U}_1, ..., \hat{U}_2]$.

### *A. Presenting $U$ with $X$*

First, let's see how to present $U$ with $X$, which implies how to estimate $U$ from $R$.

We know that:

$$[X_1, X_2] = [U_1 \oplus U_2, U_2] \tag{16}$$

which means that,

$$U_1 = X_1 \oplus X_2 \tag{17}$$

$$U_2 = X_2 \text{ and } U_2 = X_1 \oplus \hat{U}_1 \tag{18}$$

Therefore, it's clear how to decode $U_1$ from $R_1$ and $R_2$. $U_2$ can be decoded in two ways, as in Equation (18).

## References

[1] Erdal 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, July 2009.

[2] Course Instructor. Ee/stats 376a: Information theory, lecture 12, YYYY. Lecture notes.

[3] Course Instructor. Ee/stats 376a: Information theory, lecture 13, YYYY. Lecture notes.