Polar Codes

Henry Scheible, Warren Shepard

December 3, 2025

Introduction

Impact

Polar Codes were introduced by Erdal Arıkan in 2008 as the **first** channel codes to achieve the capacity of symmetric binary discrete memoryless channels and are used in 5G networks.

Introduction

Impact

Polar Codes were introduced by Erdal Arıkan in 2008 as the **first** channel codes to achieve the capacity of symmetric binary discrete memoryless channels and are used in 5G networks.

The Basic Idea

For a given channel W, channel coding is trivial when the capacity of W, denoted $C(W) \in \{0, 1\}$

1. If we can transform any channel W into a collection of extreme channels, it could be easier to communicate.
2. If the ratio of "good" channels with capacity ~ 1 to the total number of channels is $C(W)$, we can communicate with the same capacity as W.

Introduction

Impact

Polar Codes were introduced by Erdal Arıkan in 2008 as the **first** channel codes to achieve the capacity of symmetric binary discrete memoryless channels and are used in 5G networks.

The Basic Idea

For a given channel W, channel coding is trivial when the capacity of W, denoted $C(W) \in \{0, 1\}$

1. If we can transform any channel W into a collection of extreme channels, it could be easier to communicate.
2. If the ratio of "good" channels with capacity ~ 1 to the total number of channels is $C(W)$, we can communicate with the same capacity as W.

Note: when referring to an arbitrary "channel" W we mean an arbitrary symmetric binary DMC.

Agenda

Outline

Simplest Case: 2 BEC's

Recursive Construction

Channel Polarization Theorem

Code Construction

Proof of Channel Polarization Theorem

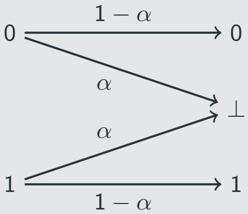
Further Results

Sending a message over two BECs

Recall:

Binary Erasure Channel (BEC)

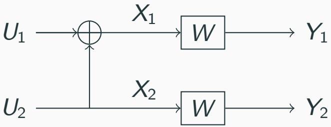
Let BEC_α be the channel with $\mathscr{X} = \{0,1\}$, $\mathscr{Y} = \{0,\perp,1\}$ and transition probabilities



for some $\alpha < 1$.

Sending a message over two BECs

For now, let $W = BEC_\alpha$ for some fixed α. Consider the compound channel W_2 sending a message in $\mathscr{X}^2$ to $\mathscr{Y}^2$ by sending each bit through W. Denote the input by the random variables X_1, X_2 and output by Y_1, Y_2.



When sending a string $u = U_1 U_2 \in \{0, 1\}$ through W_2, are we more likely to successfully decode U_1 or U_2?

Decoding & Virtual Channels

Assume that $U_1, U_2 \sim \text{Bern}(1/2)$. We will first attempt to decode U_1 knowing both Y_1 and Y_2, then if successful decode U_2 knowing Y_1, Y_2 and U_1. Call these "virtual" channels $W_2^{(1)}$ and $W_2^{(2)}$:

- $W_2^{(1)} : U_1 \mapsto Y_1 Y_2$
- $W_2^{(2)} : U_2 \mapsto Y_1 Y_2 U_1$

Conservation of Capacity

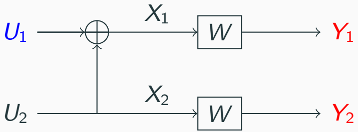
- $W_2^{(1)} : U_1 \mapsto Y_1 Y_2$
- $W_2^{(2)} : U_2 \mapsto Y_1 Y_2 U_1$

Using the Chain Rule for Mutual Information and the independence of U_1 and U_2, see that

$$\begin{aligned} C(W_2) &= I(U_1 U_2 : Y_1 Y_2) \\ &= I(U_1 : Y_1 Y_2) + I(U_2 : Y_1 Y_2 | U_1) \\ &= I(U_1 : Y_1 Y_2) + I(U_2 : Y_1 Y_2 U_1) \\ &= C\left(W_2^{(1)}\right) + C\left(W_2^{(2)}\right) \end{aligned}$$

Capacity of $W_2^{(1)}$

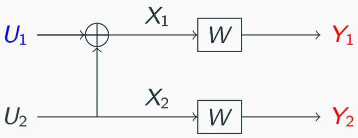
$W_2^{(1)} : U_1 \mapsto Y_1 Y_2$



$$C(W_2^{(1)}) = I(U_1 : Y_1 Y_2)$$

Capacity of $W_2^{(1)}$

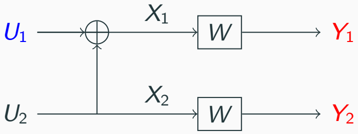
$W_2^{(1)} : U_1 \mapsto Y_1 Y_2$



$$
\begin{aligned}
C(W_2^{(1)}) &= I(U_1 : Y_1 Y_2) \\
&= H(U_1) - H(U_1 | Y_1 Y_2)
\end{aligned}
$$

Capacity of $W_2^{(1)}$

$W_2^{(1)} : U_1 \mapsto Y_1 Y_2$

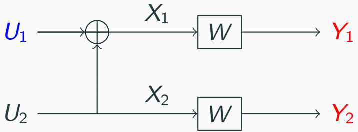


$$C(W_2^{(1)}) = I(U_1 : Y_1 Y_2)$$

$$= 1 - \sum_{(y_1, y_2) \in \{0,1,\perp\}^2} H(U_1 | Y_1 Y_2 = y_1 y_2) \mathbb{P}\{Y_1 Y_2 = y_1 y_2\}$$

Capacity of $W_2^{(1)}$

$W_2^{(1)} : U_1 \mapsto Y_1 Y_2$

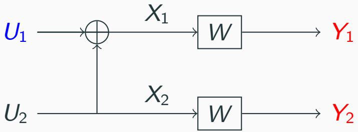


Note that $U_1 U_2 \sim \text{Bern}(1/2)^{\otimes 2}$ implies that $X_1 X_2 \sim \text{Bern}(1/2)^{\otimes 2}$, so $Y_1, Y_2 \sim p$, where

- $p(0) = 1/2(1-\alpha)$
- $p(\perp) = \alpha$
- $p(1) = 1/2(1-\alpha)$

Capacity of $W_2^{(1)}$

$W_2^{(1)} : U_1 \mapsto Y_1 Y_2$



Note that $U_1 U_2 \sim \text{Bern}(1/2)^{\otimes 2}$ implies that $X_1 X_2 \sim \text{Bern}(1/2)^{\otimes 2}$, so $Y_1, Y_2 \sim p$, where

- $p(0) = 1/2(1-\alpha)$
- $p(\perp) = \alpha$
- $p(1) = 1/2(1-\alpha)$

If $Y_1, Y_2 \neq \perp$ (w.p. $(1-\alpha)^2$), then

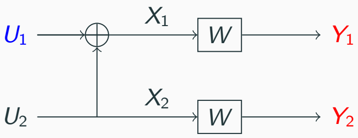
$$H(U_1) = H_2(0) = 0$$

If $Y_1 = \perp$ or $Y_2 = \perp$ (w.p. $1-(1-\alpha)^2$), then

$$H(U_1) = H_2(1/2) = 1$$

Capacity of $W_2^{(1)}$

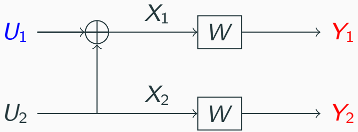
$W_2^{(1)} : U_1 \mapsto Y_1 Y_2$



$$
\begin{aligned}
C(W_2^{(1)}) &= I(U_1 : Y_1 Y_2) \\
&= H(U_1) - H(U_1 | Y_1 Y_2) \\
&= 1 - \sum_{(y_1, y_2) \in \{0,1,\perp\}^2} H(U_1 | Y_1 Y_2 = y_1 y_2) \mathbb{P}\{Y_1 Y_2 = y_1 y_2\}
\end{aligned}
$$

Capacity of $W_2^{(1)}$

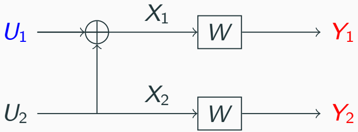
$W_2^{(1)} : U_1 \mapsto Y_1 Y_2$



$$
\begin{aligned}
C(W_2^{(1)}) &= I(U_1 : Y_1 Y_2) \\
&= H(U_1) - H(U_1 | Y_1 Y_2) \\
&= 1 - \sum_{(y_1, y_2) \in \{0,1,\perp\}^2} H(U_1 | Y_1 Y_2 = y_1 y_2) \mathbb{P}\{Y_1 Y_2 = y_1 y_2\} \\
&= 1 - (1 - (1-\alpha)^2) = (1-\alpha)^2
\end{aligned}
$$

Capacity of $W_2^{(1)}$

$W_2^{(1)} : U_1 \mapsto Y_1 Y_2$



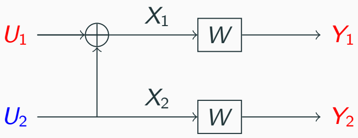
$$
\begin{aligned}
C(W_2^{(1)}) &= I(U_1 : Y_1 Y_2) \\
&= H(U_1) - H(U_1 | Y_1 Y_2) \\
&= 1 - \sum_{(y_1, y_2) \in \{0,1,\perp\}^2} H(U_1 | Y_1 Y_2 = y_1 y_2) \mathbb{P}\{Y_1 Y_2 = y_1 y_2\} \\
&= 1 - (1 - (1-\alpha)^2) = (1-\alpha)^2
\end{aligned}
$$

Remark

$W_2^{(1)}$ is effectively $BEC_{1-(1-\alpha)^2}$

Capacity of $W_2^{(2)}$

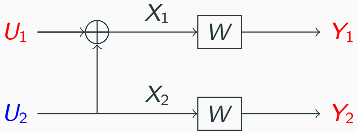
$W_2^{(1)} : U_2 \mapsto Y_1 Y_2 U_1$



$$C(W_2^{(2)}) = I(U_2 : Y_1 Y_2 U_1)$$

Capacity of $W_2^{(2)}$

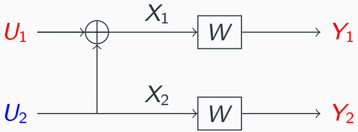
$W_2^{(1)} : U_2 \mapsto Y_1 Y_2 U_1$



$$
\begin{aligned}
C(W_2^{(2)}) &= I(U_2 : Y_1 Y_2 U_1) \\
&= H(U_2) - H(U_2 | Y_1 Y_2 U_1)
\end{aligned}
$$

Capacity of $W_2^{(2)}$

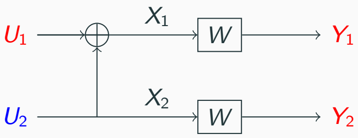
$W_2^{(1)} : U_2 \mapsto Y_1 Y_2 U_1$



$$
\begin{aligned}
C(W_2^{(2)}) &= I(U_2 : Y_1 Y_2 U_1) \\
&= H(U_2) - H(U_2 | Y_1 Y_2 U_1) \\
&= 1 - \sum_{(y_1, y_2, u_1) \in \{0,1,\perp\}^2 \times \{0,1\}} H(U_2 | Y_1 Y_2 U_1 = y_1 y_2 u_1) \mathbb{P}\{Y_1 Y_2 U_1 = y_1 y_2 u_1\}
\end{aligned}
$$

Capacity of $W_2^{(2)}$

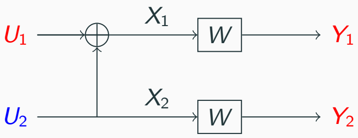
$W_2^{(1)} : U_2 \mapsto Y_1 Y_2 U_1$



With U_1 fixed, X_1 and X_2 are both deterministic functions of U_2.

Capacity of $W_2^{(2)}$

$W_2^{(1)} : U_2 \mapsto Y_1 Y_2 U_1$



With U_1 fixed, X_1 and X_2 are both deterministic functions of U_2.

If $Y_1 \neq \perp$ or $Y_2 \neq \perp$ (w.p. $1 - \alpha^2$), then

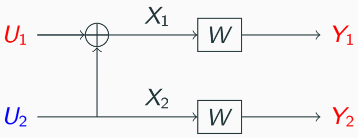
$$H(U_2) = H_2(0) = 0$$

If $Y_1 = Y_2 = \perp$ (w.p. α^2), then

$$H(U_2) = H_2(1/2) = 1$$

Capacity of $W_2^{(2)}$

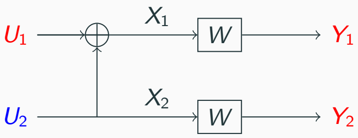
$W_2^{(1)} : U_2 \mapsto Y_1 Y_2 U_1$



$$
\begin{aligned}
C(W_2^{(2)}) &= I(U_2 : Y_1 Y_2 U_1) \\
&= H(U_2) - H(U_2 | Y_1 Y_2 U_1) \\
&= 1 - \sum_{(y_1, y_2, u_1) \in \{0,1,\perp\}^2 \times \{0,1\}} H(U_2 | Y_1 Y_2 U_1 = y_1 y_2 u_1) \mathbb{P}\{Y_1 Y_2 U_1 = y_1 y_2 u_1\}
\end{aligned}
$$

Capacity of $W_2^{(2)}$

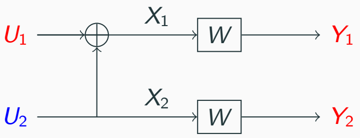
$W_2^{(1)} : U_2 \mapsto Y_1 Y_2 U_1$



$$
\begin{aligned}
C(W_2^{(2)}) &= I(U_2 : Y_1 Y_2 U_1) \\
&= H(U_2) - H(U_2 | Y_1 Y_2 U_1) \\
&= 1 - \sum_{(y_1, y_2, u_1) \in \{0,1,\perp\}^2 \times \{0,1\}} H(U_2 | Y_1 Y_2 U_1 = y_1 y_2 u_1) \mathbb{P}\{Y_1 Y_2 U_1 = y_1 y_2 u_1\} \\
&= 1 - \alpha^2
\end{aligned}
$$

Capacity of $W_2^{(2)}$

$W_2^{(1)} : U_2 \mapsto Y_1 Y_2 U_1$



$$
\begin{aligned}
C(W_2^{(2)}) &= I(U_2 : Y_1 Y_2 U_1) \\
&= H(U_2) - H(U_2 | Y_1 Y_2 U_1) \\
&= 1 - \sum_{(y_1, y_2, u_1) \in \{0,1,\perp\}^2 \times \{0,1\}} H(U_2 | Y_1 Y_2 U_1 = y_1 y_2 u_1) \mathbb{P}\{Y_1 Y_2 U_1 = y_1 y_2 u_1\} \\
&= 1 - \alpha^2
\end{aligned}
$$

Remark

$W_2^{(2)}$ is effectively BEC_{α^2}

Summary of $W_2^{(1)}$ and $W_2^{(2)}$

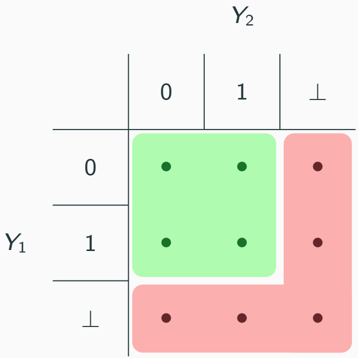


Figure 1: $W_2^{(1)}$

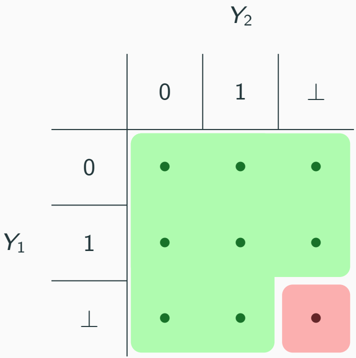


Figure 2: $W_2^{(2)}$

BEC_{α^2}
Erasure occurs if *either* channel erases

$BEC_{1-(1-\alpha)^2}$
Erasure occurs if *both* channels erase

Outline

Simplest Case: 2 BEC's

Recursive Construction

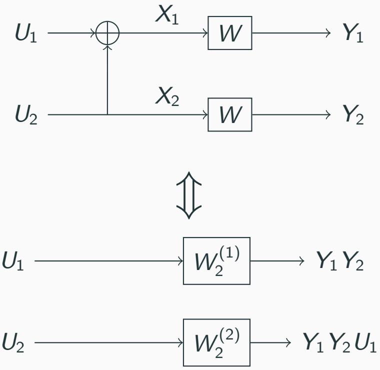
Channel Polarization Theorem

Code Construction

Proof of Channel Polarization Theorem

Further Results

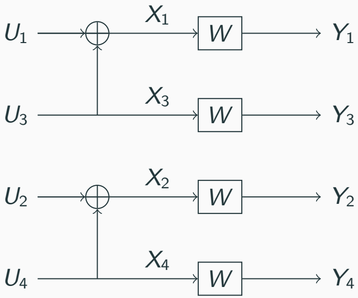
Channel Splitting



Recursive Construction of Polarizing Channels

Idea

We can use $W_2^{(1)}$ and $W_2^{(2)}$ to recursively construct more polarizing channels the same way we constructed $W_2^{(1)}$ and $W_2^{(2)}$ from W.



Recursive Construction of Polarizing Channels

Idea

We can use $W_2^{(1)}$ and $W_2^{(2)}$ to recursively construct more polarizing channels the same way we constructed $W_2^{(1)}$ and $W_2^{(2)}$ from W.

$$U_1 \longrightarrow \boxed{W_2^{(1)}} \longrightarrow Y_1 Y_3$$

$$U_3 \longrightarrow \boxed{W_2^{(2)}} \longrightarrow Y_1 Y_3 U_1$$

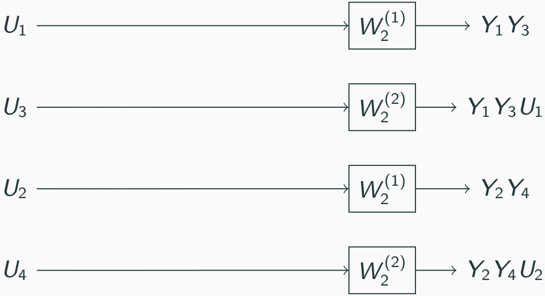
$$U_2 \longrightarrow \boxed{W_2^{(1)}} \longrightarrow Y_2 Y_4$$

$$U_4 \longrightarrow \boxed{W_2^{(2)}} \longrightarrow Y_2 Y_4 U_2$$

Recursive Construction of Polarizing Channels

Idea

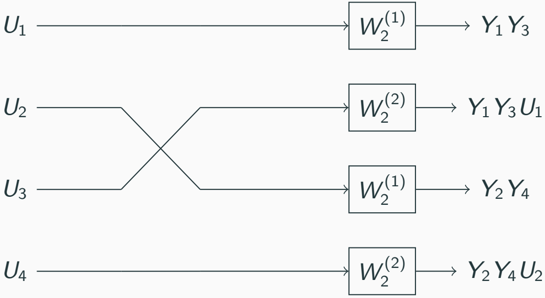
We can use $W_2^{(1)}$ and $W_2^{(2)}$ to recursively construct more polarizing channels the same way we constructed $W_2^{(1)}$ and $W_2^{(2)}$ from W.



Recursive Construction of Polarizing Channels

Idea

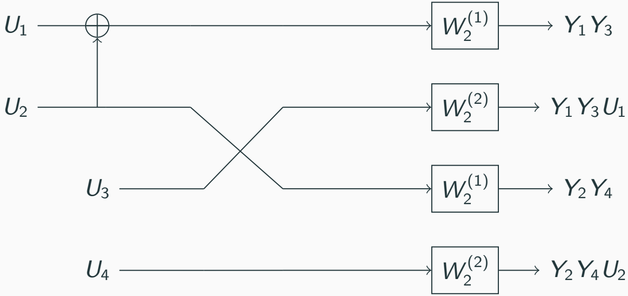
We can use $W_2^{(1)}$ and $W_2^{(2)}$ to recursively construct more polarizing channels the same way we constructed $W_2^{(1)}$ and $W_2^{(2)}$ from W.



Recursive Construction of Polarizing Channels

Idea

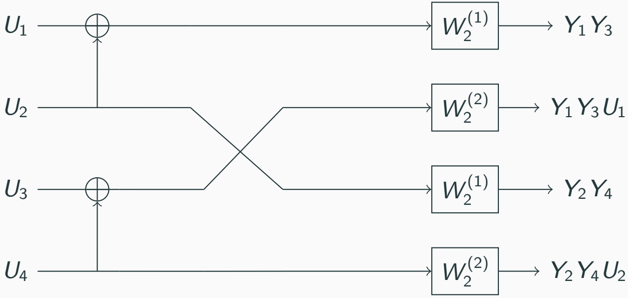
We can use $W_2^{(1)}$ and $W_2^{(2)}$ to recursively construct more polarizing channels the same way we constructed $W_2^{(1)}$ and $W_2^{(2)}$ from W.



Recursive Construction of Polarizing Channels

Idea

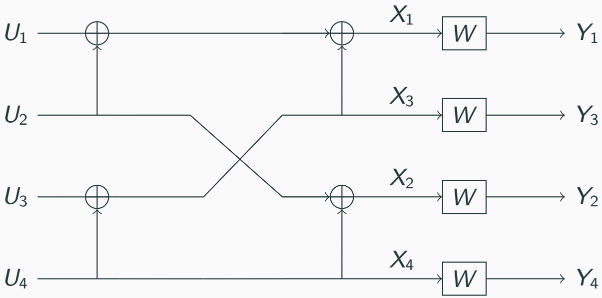
We can use $W_2^{(1)}$ and $W_2^{(2)}$ to recursively construct more polarizing channels the same way we constructed $W_2^{(1)}$ and $W_2^{(2)}$ from W.



Recursive Construction of Polarizing Channels

Idea

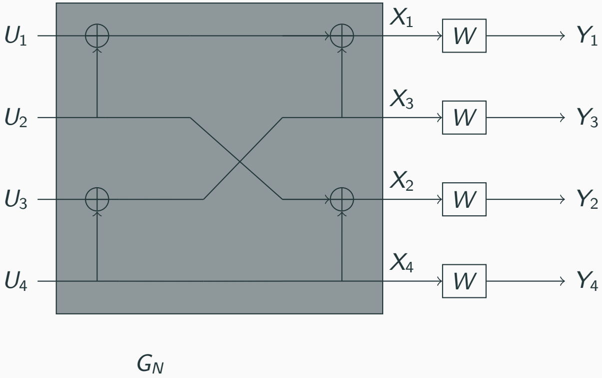
We can use $W_2^{(1)}$ and $W_2^{(2)}$ to recursively construct more polarizing channels the same way we constructed $W_2^{(1)}$ and $W_2^{(2)}$ from W.



Recursive Construction of Polarizing Channels

Idea

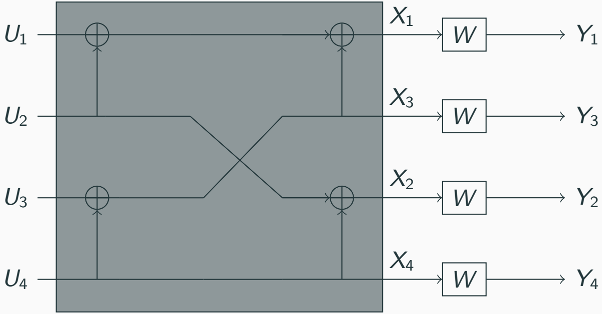
We can use $W_2^{(1)}$ and $W_2^{(2)}$ to recursively construct more polarizing channels the same way we constructed $W_2^{(1)}$ and $W_2^{(2)}$ from W.



Recursive Construction of Polarizing Channels

Idea

We can use $W_2^{(1)}$ and $W_2^{(2)}$ to recursively construct more polarizing channels the same way we constructed $W_2^{(1)}$ and $W_2^{(2)}$ from W.



$$G_N = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}^{\otimes 2}$$

Final Virtual Channels

From this process, we get four virtual channels from our four initial channels W:

- $W_4^{(1)} : U_1 \to Y_1 \cdots Y_4$
- $W_4^{(2)} : U_1 \to Y_1 \cdots Y_4 U_1$
- $W_4^{(3)} : U_1 \to Y_1 \cdots Y_4 U_1 U_2$
- $W_4^{(i)} : U_1 \to Y_1 \cdots Y_4 U_1 U_2 U_3$

Final Virtual Channels (General Case)

From this process, we get N virtual channels from our N initial channels W:

- $W_N^{(1)} : U_1 \to Y_1 \cdots Y_N$
- $W_N^{(2)} : U_1 \to Y_1 \cdots Y_N U_1$
- $\cdots$
- $W_N^{(i)} : U_1 \to Y_1 \cdots Y_N U_1 \cdots U_{i-1}$

Outline

Channel Polarization Theorem

Formalizing channel polarization:

Theorem: For any channel W and $\delta \in [0,1]$, consider the channels $\left\{W_N^{(i)}\right\}$ where $N = 2^n$ is the number of channels for an integer n. As $n \to \infty$,

1. the fraction of indices $i \in [N]$ such that $C(W_N^{(i)}) \in (1-\delta, 1]$ approaches $C(W)$
2. the fraction of indices i such that $C(W_N^{(i)}) \in [0, \delta)$ approaches $1 - C(W)$

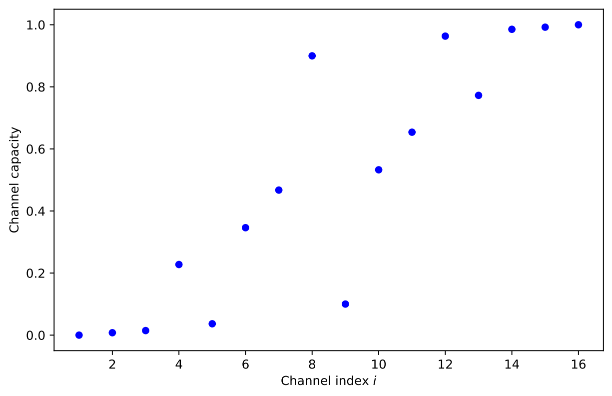
where $C(W)$ denotes the symmetric capacity of W (same as standard capacity for BEC).

Polarization Theorem Demonstration

Setup

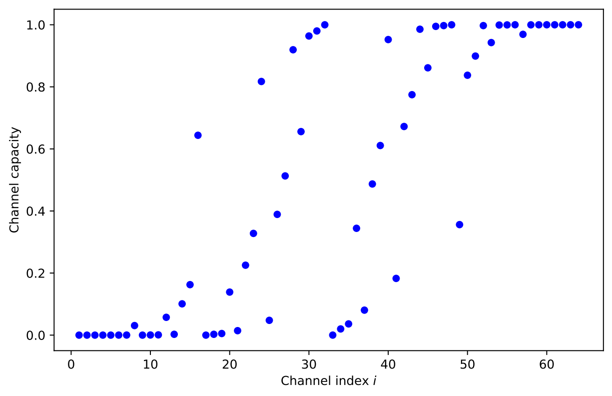
Lets revisit our example of a Binary Erasure Channel with erasure probability $\alpha < 1$. For simplicity set $\alpha = \frac{1}{2}$. Now we want to (empirically) show that as *n* increases, the resulting channels polarize.

Polarization Theorem Demonstration



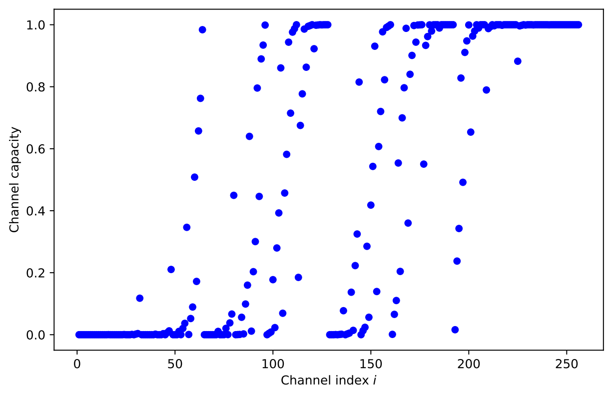
n=4, $\alpha = 0.5$

Polarization Theorem Demonstration



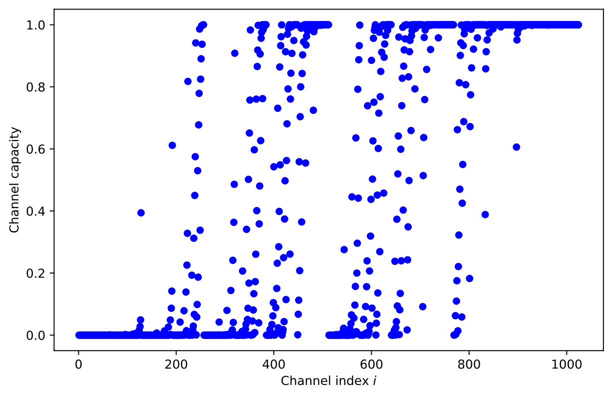
n=4, $\alpha = 0.5$

Polarization Theorem Demonstration



n=4, $\alpha = 0.5$

Polarization Theorem Demonstration



n=4, $\alpha = 0.5$

Theorem 1 Demonstration

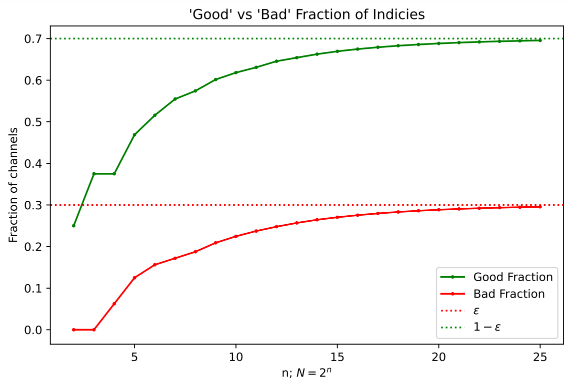
Setup

Lets revisit our example of a Binary Erasure Channel with erasure probability $\alpha < 1$. For simplicity set $\alpha = \frac{3}{10}$. Note $C = 1 - \alpha = \frac{7}{10}$, so we expect the fraction of "good" channels to approach $\frac{7}{10}$ for $\delta = 0.05$.

Theorem 1 Demonstration

Setup

Lets revisit our example of a Binary Erasure Channel with erasure probability $\alpha < 1$. For simplicity set $\alpha = \frac{3}{10}$. Note $C = 1 - \alpha = \frac{7}{10}$, so we expect the fraction of "good" channels to approach $\frac{7}{10}$ for $\delta = 0.05$.



Outline

Constructing a Polar Code

Given a symmetric DMC W^N which is split into $W_n^{(1)}, ..., W_n^{(n)}$ polarized channels, we define a polar code with rate k/n by picking k channels with $I(W_n^{(i)})$ maximized.
Alice and Bob agree in advance on the values to be sent on the $n-k$ remaining channels, then Alice sends her message on those k channels.
Following the theorem, as $n \to \infty$ Alice will be able to pick a capacity-achieving rate with polarized channel capacity approaching 1.

Successive Cancellation Decoder

To *actually* split W^N into $W_N^{(1)}, ..., W_N^{(N)}$, we must use a particular decoding strategy: *successive cancellation decoding*:

For each $i \in 1, ..., N$:

- If i is in a meaningful channel: Decode $\hat{U}_i$ based on maximum likelihood from $Y_1, ..., Y_N, \hat{U}_1, ..., \hat{U_{i-1}}$
    - Compare $W_N^{(i)}(Y_1 \cdots Y_N U_{1\,i-1}|0)$ with $W_N^{(i)}(Y_1 \cdots Y_N U_{1\,i-1}|1)$
- If i is in a meaningless channel: set $\hat{U}_i = U_i$

Outline

Simplest Case: 2 BEC's

Recursive Construction

Channel Polarization Theorem

Code Construction

Proof of Channel Polarization Theorem

Further Results

Proof of Main Theorem: Goal

Theorem: For any Binary DMC W and $\delta \in [0,1]$, consider the channels $\left\{W_N^{(i)}\right\}$ where $N = 2^n$ is the number of channels for an integer n. As $n \to \infty$,

1. the fraction of indices $i \in [N]$ such that $C(W_N^{(i)}) \in (1-\delta, 1]$ approaches $C(W)$
2. the fraction of indices i such that $C(W_N^{(i)}) \in [0, \delta)$ approaches $1 - C(W)$

where $C(W)$ denotes the symmetric capacity of W (same as standard capacity for BEC).

Proof of Main Theorem: Preliminaries

Let $Z(W)$ be the "reliability" of a channel, defined as

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

Proof of Main Theorem: Preliminaries

Let $Z(W)$ be the "reliability" of a channel, defined as

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

The symmetric capacity, $C(W)$, of a channel is defined as

$$C(W) = \sum_{x \in \mathscr{X}} \sum_{y \in \mathscr{Y}} \frac{1}{2} W(y \mid x) \log \frac{W(y \mid x)}{\frac{1}{2} W(y \mid 0) + \frac{1}{2} W(y \mid 1)}.$$

Proof of Main Theorem: Preliminaries

Let $Z(W)$ be the "reliability" of a channel, defined as

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

The symmetric capacity, $C(W)$, of a channel is defined as

$$C(W) = \sum_{x \in \mathscr{X}} \sum_{y \in \mathscr{Y}} \frac{1}{2} W(y \mid x) \log \frac{W(y \mid x)}{\frac{1}{2} W(y \mid 0) + \frac{1}{2} W(y \mid 1)}.$$

Consider the binary tree representation of a polar code construction. Let $\{K_n\} = \{K_n \mid n \geq 0\}$ be a random walk through the tree beginning at the root ($K_0 = W$) and ending at a leaf.

Proof of Main Theorem: Preliminaries

Let $Z(W)$ be the "reliability" of a channel, defined as

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

The symmetric capacity, $C(W)$, of a channel is defined as

$$C(W) = \sum_{x \in \mathscr{X}} \sum_{y \in \mathscr{Y}} \frac{1}{2} W(y \mid x) \log \frac{W(y \mid x)}{\frac{1}{2} W(y \mid 0) + \frac{1}{2} W(y \mid 1)}.$$

Consider the binary tree representation of a polar code construction. Let $\{K_n\} = \{K_n \mid n \geq 0\}$ be a random walk through the tree beginning at the root $(K_0 = W)$ and ending at a leaf.

To keep track of all the rate and reliability parameters, let $\{C_n\}$ and $\{Z_n\}$ be the capacity and reliability sequences corresponding to $\{K_n\}$.

Proof of Main Theorem: Preliminaries

Let $Z(W)$ be the "reliability" of a channel, defined as

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

The symmetric capacity, $C(W)$, of a channel is defined as

$$C(W) = \sum_{x \in \mathscr{X}} \sum_{y \in \mathscr{Y}} \frac{1}{2} W(y \mid x) \log \frac{W(y \mid x)}{\frac{1}{2} W(y \mid 0) + \frac{1}{2} W(y \mid 1)}.$$

Consider the binary tree representation of a polar code construction. Let $\{K_n\} = \{K_n \mid n \geq 0\}$ be a random walk through the tree beginning at the root ($K_0 = W$) and ending at a leaf.

To keep track of all the rate and reliability parameters, let $\{C_n\}$ and $\{Z_n\}$ be the capacity and reliability sequences corresponding to $\{K_n\}$.

$\mathscr{F}$ is the Borel Field generated by the cylinder sets $S(b_1, \dots b_n)$ where $b_1, \dots b_n$ is a binary sequence denoting the traversal of the binary tree representation.

Proof of Main Theorem: Lemma 1

Lemma 1

$\{C_n, F_n\}$ is a martingale (a sequence of RVs with the conditional expectation equal to the current value), i.e.*

1. $\mathbf{Exp}[|C_n|] < \infty$
2. $c_n = \mathbf{Exp}[C_{n+1} \mid C_n = c_n]$

Notably, $\{C_n\}$ converges almost always to C_∞ with $\mathbf{Exp}[C_\infty] = C_0$.

Proof of Main Theorem: Lemma 1

Lemma 1

$\{C_n, F_n\}$ is a martingale (a sequence of RVs with the conditional expectation equal to the current value), i.e.*

1. $\mathbf{Exp}[|C_n|] < \infty$
2. $c_n = \mathbf{Exp}[C_{n+1} \mid C_n = c_n]$

Notably, $\{C_n\}$ converges almost always to C_∞ with $\mathbf{Exp}[C_\infty] = C_0$.

Proof. (1) is true based on the fact that $0 \leq C_n \leq 1$. To show (2), consider a binary sequence $(b_1, \ldots, b_n) \in \{0,1\}^n$ and let $W_{b_1 \ldots b_n}$ be the channel obtained by traversing the tree representation based on the binary sequence.

$$\begin{aligned}\mathbf{Exp}[C_{n+1} \mid S(b_1, \ldots, b_n)] &= \frac{1}{2} C(W_{b_1 \ldots b_n 0}) + \frac{1}{2} C(W_{b_1 \ldots b_n 1}) \\ &= C(W_{b_1 \ldots b_n})\end{aligned}$$

Proof of Main Theorem: Lemma 1

Lemma 1

$\{C_n, F_n\}$ is a martingale (a sequence of RVs with the conditional expectation equal to the current value), i.e.*

1. $\mathbf{Exp}[|C_n|] < \infty$
2. $c_n = \mathbf{Exp}[C_{n+1} \mid C_n = c_n]$

Notably, $\{C_n\}$ converges almost always to C_∞ with $\mathbf{Exp}[C_\infty] = C_0$.

Proof. (1) is true based on the fact that $0 \leq C_n \leq 1$. To show (2), consider a binary sequence $(b_1, \ldots, b_n) \in \{0,1\}^n$ and let $W_{b_1 \ldots b_n}$ be the channel obtained by traversing the tree representation based on the binary sequence.

$$\begin{aligned} \mathbf{Exp}[C_{n+1} \mid S(b_1, \ldots, b_n)] &= \frac{1}{2} C(W_{b_1 \ldots b_n 0}) + \frac{1}{2} C(W_{b_1 \ldots b_n 1}) \\ &= C(W_{b_1 \ldots b_n}) \end{aligned}$$

Since (C_n, F_n) is a martingale, it follows that $\{C_n\}$ converges almost always (explicitly, a set of points with measure zero can be excluded from convergence) to some random variable C_∞ such that $\mathbf{Exp}[C_\infty] = C_0$.

Proof of Main Theorem: Lemma 2

Lemma 2

$\{Z_n, F_n\}$ is a super-martingale (conditional expectation of next value is bounded by current), i.e.*

1. $\mathbf{Exp}[|Z_n|] < \infty$
2. $z_n \geq \mathbf{Exp}[Z_{n+1} \mid Z_n = z_n]$

Notably, $\{Z_n\}$ converges to Z_∞ which almost always takes value $\{0, 1\}$.

Proof of Main Theorem: Lemma 2

Lemma 2

$\{Z_n, F_n\}$ is a super-martingale (conditional expectation of next value is bounded by current), i.e.*

1. $\mathbf{Exp}[|Z_n|] < \infty$
2. $z_n \geq \mathbf{Exp}[Z_{n+1} \mid Z_n = z_n]$

Notably, $\{Z_n\}$ converges to Z_∞ which almost always takes value $\{0, 1\}$.

Proof. (1) is satisfied because $Z(W)$ takes on values in $[0, 1]$. To prove (2), we follow a same structure as for I_n. Again, consider a binary $b_1, \ldots, b_n$. Then,

$$\begin{aligned}\mathbf{Exp}[Z_{n+1} \mid S(b_1, \ldots, b_n)] &= \frac{1}{2} Z(W_{b_1 \ldots b_n 0}) + \frac{1}{2} Z(W_{b_1 \ldots b_n 1}) \\ &\leq Z(W_{b_1 \ldots b_n})\end{aligned}$$

where $W_{b_1 \ldots b_n}$ is the value of Z_n on $b_1, \ldots b_n$

Proof of Main Theorem: Lemma 2

Since $\{Z_n, F_n\}$ is a super-martingale, Z_n converges to a random variable Z_∞ such that $\mathbf{Exp}[|Z_n - Z_\infty|] \to 0$. Note $Z_{n+1} = Z_n^2$ with probability $1/2$, so

$$\mathbf{Exp}[|Z_{n+1} - Z_n|] \geq \frac{1}{2}\mathbf{Exp}[Z_n^2 - Z_n] = \frac{1}{2}\mathbf{Exp}[Z_n(1 - Z_n)] \geq 0.$$

Proof of Main Theorem: Lemma 2

Since $\{Z_n, F_n\}$ is a super-martingale, Z_n converges to a random variable Z_∞ such that $\mathbf{Exp}[|Z_n - Z_\infty|] \to 0$. Note $Z_{n+1} = Z_n^2$ with probability $1/2$, so

$$\mathbf{Exp}[|Z_{n+1} - Z_n|] \geq \frac{1}{2}\mathbf{Exp}[Z_n^2 - Z_n] = \frac{1}{2}\mathbf{Exp}[Z_n(1 - Z_n)] \geq 0.$$

Therefore,

$$\mathbf{Exp}[Z_n(1 - Z_n)] \to 0 \implies \mathbf{Exp}[Z_\infty(1 - Z_\infty)] \to 0.$$

For the above to hold, Z_∞ must almost always be 0 or 1.

Proof of Main Theorem: Conclusion

We have that Z_∞ converges to 0 or 1 almost always. It is a fact [proved in appendix of paper] that:

$$C(W) \geq \log \frac{2}{1+Z(W)} \quad \text{and} \quad C(W) \leq \sqrt{1-Z(W)^2}.$$

Proof of Main Theorem: Conclusion

We have that Z_∞ converges to 0 or 1 almost always. It is a fact [proved in appendix of paper] that:

$$C(W) \geq \log \frac{2}{1 + Z(W)} \quad \text{and} \quad C(W) \leq \sqrt{1 - Z(W)^2}.$$

It follows that $C_\infty = 1 - Z_\infty$. Since $\mathbf{Exp}[C_\infty] = C_0$, we get that C_∞ converges to 1 with probability C_0 and converges to 0 with probability $1 - C_0$.

Proof of Main Theorem: Conclusion

We have that Z_∞ converges to 0 or 1 almost always. It is a fact [proved in appendix of paper] that:

$$C(W) \geq \log \frac{2}{1+Z(W)} \quad \text{and} \quad C(W) \leq \sqrt{1-Z(W)^2}.$$

It follows that $C_\infty = 1 - Z_\infty$. Since $\mathbf{Exp}[C_\infty] = C_0$, we get that C_∞ converges to 1 with probability C_0 and converges to 0 with probability $1 - C_0$.

Theorem: For any channel W and $\delta \in [0,1]$, consider the channels $\left\{W_N^{(i)}\right\}$ where $N = 2^n$ is the number of channels for an integer n. As $n \to \infty$,

1. the fraction of indices $i \in [N]$ such that $C(W_N^{(i)}) \in (1-\delta, 1]$ approaches $C(W)$
2. the fraction of indices i such that $C(W_N^{(i)}) \in [0, \delta)$ approaches $1 - C(W)$

Outline

Simplest Case: 2 BEC's

Recursive Construction

Channel Polarization Theorem

Code Construction

Proof of Channel Polarization Theorem

Further Results

Bound on the Rate of Polarization

Theorem: For any Binary DMC W with $C(W) > 0$ and any fixed $R < C(W)$, there exists a sequence of sets $\mathscr{A}_N \subset \{1, ..., N\}, N \in \{1, 2, ..., 2^n, ...\}$, such that $|\mathscr{A}_N| \geq NR$ and $Z(W_N^{(i)}) \leq O(N^{-5/4})$ for all $i \in \mathscr{A}_n s$

Recall that

$$C(W) \geq \log \frac{2}{1 + Z(W)}$$

Recent Result: List Decoding for Polar Codes (2015)

Our current channel polarization result relies on *successive cancellation decoding*, which is significantly weaker than *maximum likelihood decoding*

A recent advance proposes keeping L candidate strings at all times then picking the most likely one at the end of the decoding process, a decoding strategy which is between the other two options.

This makes Polar Codes actually useful for small to medium block lengths.

References i

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, July 2009.

E. Arikan.
Polar codes i (youtube lecture) - erdal arikan, 2015.
Accessed: 6-Mar-2025.

E. Arikan.
Polar codes ii (youtube lecture) - erdal arikan, 2015.
Accessed: 6-Mar-2025.

E. Arikan.
Polar codes iii (youtube lecture) - erdal arikan, 2015.
Accessed: 6-Mar-2025.

H. D. Pfister.
A brief introduction to polar codes, 2017.
Notes for Introduction to Error-Correcting Codes, Accessed: 6-Mar-2025.

References ii

I. Tal and A. Vardy.
List decoding of polar codes.
IEEE Transactions on Information Theory, 61(5):2213–2226, 2015.