Polar Codes

Aayush Rajesh, Ronil Mandavia

Department of Electrical Engineering
IIT Bombay

Outline

Overview

- Polar Codes were introduced by Erdal Arıkan in [1], and are constructed by exploiting a phenomenon known as **channel polarization**.

Overview

- Polar Codes were introduced by Erdal Arıkan in [1], and are constructed by exploiting a phenomenon known as **channel polarization**.
- It can be shown that these codes achieve Shannon's Capacity.

Overview

- Polar Codes were introduced by Erdal Arıkan in [1], and are constructed by exploiting a phenomenon known as **channel polarization**.
- It can be shown that these codes achieve Shannon's Capacity.
- We shall look to motivate the polarization phenomenon to understand how these codes achieve capacity. Efficient encoding and decoding procedures shall also be covered.

Information Theoretic Measures

- **Entropy:** Entropy measures the average uncertainty in a random variable.

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x)$$

Information Theoretic Measures

- **Entropy:** Entropy measures the average uncertainty in a random variable.

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x)$$

$$H(Y|X) = \sum_{x \in \mathcal{X}} p(x) H(Y|X = x)$$

Information Theoretic Measures

- **Entropy:** Entropy measures the average uncertainty in a random variable.

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x)$$

$$H(Y|X) = \sum_{x \in \mathcal{X}} p(x) H(Y|X = x)$$

- **Mutual Information:** The reduction in uncertainty in one random variable due to knowledge of another.

$$I(X;Y) = H(X) - H(X|Y)$$

Information Theoretic Measures

- **Entropy:** Entropy measures the average uncertainty in a random variable.

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x)$$

$$H(Y|X) = \sum_{x \in \mathcal{X}} p(x) H(Y|X = x)$$

- **Mutual Information:** The reduction in uncertainty in one random variable due to knowledge of another.

$$I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

Information Theoretic Measures

- **Entropy:** Entropy measures the average uncertainty in a random variable.

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x)$$

$$H(Y|X) = \sum_{x \in \mathcal{X}} p(x) H(Y|X = x)$$

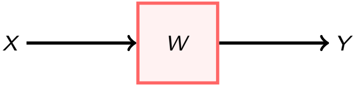
- **Mutual Information:** The reduction in uncertainty in one random variable due to knowledge of another.

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

- **Chain Rule:**

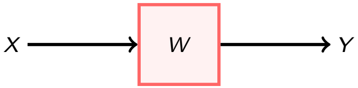
$$I(X,Y;Z) = I(X;Z) + I(Y;Z|X)$$

The Channel



- Channel $W : \mathcal{X} \to \mathcal{Y}$, characterized by **transition probabilities** $W(y|x)$.

The Channel



- Channel $W : \mathcal{X} \to \mathcal{Y}$, characterized by **transition probabilities** $W(y|x)$.
- The maximum rate of reliable communication over this B-DMC[1] is the **capacity** and is given by

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

[1] Binary-Discrete Memoryless Channel

The Channel

$$X \longrightarrow \boxed{W} \longrightarrow Y$$

- Channel $W : \mathcal{X} \rightarrow \mathcal{Y}$, characterized by **transition probabilities** $W(y|x)$.
- The maximum rate of reliable communication over this B-DMC[1] is the **capacity** and is given by

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

- There are two classes for which analysis of communication is trivial
  - Perfect Channels: $C = 1$
  - Useless Channels: $C = 0$

[1] Binary-Discrete Memoryless Channel

Channel Capacity

A simple analysis of channel capacity can be undertaken for the BEC(p) channel. We have $W(0|0) = W(1|1) = 1 - p$ and $W(?|0) = W(?|1) = p$. The mutual information of the channel can be evaluated as

Channel Capacity

A simple analysis of channel capacity can be undertaken for the BEC(p) channel. We have $W(0|0) = W(1|1) = 1 - p$ and $W(?|0) = W(?|1) = p$. The mutual information of the channel can be evaluated as

$$I(X;Y) = H(X) - H(X|Y)$$

Channel Capacity

A simple analysis of channel capacity can be undertaken for the BEC(p) channel. We have $W(0|0) = W(1|1) = 1 - p$ and $W(?|0) = W(?|1) = p$. The mutual information of the channel can be evaluated as

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \\ &= H(X) - p_Y(?)H(X|?) - p_Y(0)H(X|0) - p_Y(1)H(X|1) \end{aligned}$$

Channel Capacity

A simple analysis of channel capacity can be undertaken for the BEC(p) channel. We have $W(0|0) = W(1|1) = 1 - p$ and $W(?|0) = W(?|1) = p$. The mutual information of the channel can be evaluated as

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \\ &= H(X) - p_Y(?)H(X|?) - p_Y(0)H(X|0) - p_Y(1)H(X|1) \\ &= H(X) - pH(X) - 0 - 0 \end{aligned}$$

Channel Capacity

A simple analysis of channel capacity can be undertaken for the BEC(p) channel. We have $W(0|0) = W(1|1) = 1 - p$ and $W(?|0) = W(?|1) = p$. The mutual information of the channel can be evaluated as

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \\ &= H(X) - p_Y(?)H(X|?) - p_Y(0)H(X|0) - p_Y(1)H(X|1) \\ &= (1-p)H(X) \end{aligned}$$

Channel Capacity

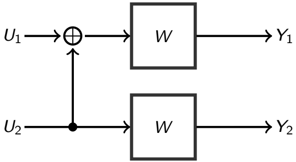
A simple analysis of channel capacity can be undertaken for the BEC(p) channel. We have $W(0|0) = W(1|1) = 1 - p$ and $W(?|0) = W(?|1) = p$. The mutual information of the channel can be evaluated as

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \\ &= H(X) - p_Y(?)H(X|?) - p_Y(0)H(X|0) - p_Y(1)H(X|1) \\ &= (1-p)H(X) \end{aligned}$$

This quantity is maximized when $X \sim \text{Ber}\left(\frac{1}{2}\right)$, giving us $C = 1 - p$ for a BEC(p) channel. Through a similar analysis, we can obtain that $C = 1 - H(p)$ for a BSC(p) channel.

Combining Channels

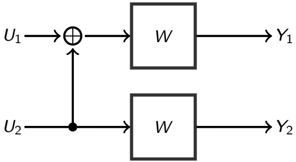
- We can denote $X_1 = U_1 \oplus U_2$ and $X_2 = U_2$. There is an invertible transformation between (U_1, U_2) and (X_1, X_2).



Combining Channels

- We can denote $X_1 = U_1 \oplus U_2$ and $X_2 = U_2$. There is an invertible transformation between (U_1, U_2) and (X_1, X_2).
- The capacity of this joint ensemble is

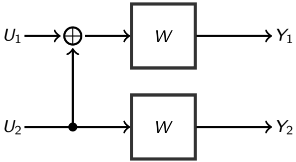
$$2C(W) = I(X_1 X_2; Y_1 Y_2) = I(U_1 U_2; Y_1 Y_2)$$



Combining Channels

- We can denote $X_1 = U_1 \oplus U_2$ and $X_2 = U_2$. There is an invertible transformation between (U_1, U_2) and (X_1, X_2).
- The capacity of this joint ensemble is

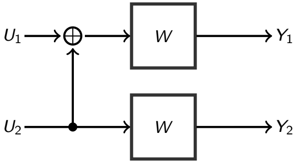
$$\begin{aligned} 2C(W) &= I(X_1 X_2; Y_1 Y_2) = I(U_1 U_2; Y_1 Y_2) \\ &= I(U_1; Y_1 Y_2) + I(U_2; Y_1 Y_2 | U_1) \end{aligned}$$



Combining Channels

- We can denote $X_1 = U_1 \oplus U_2$ and $X_2 = U_2$. There is an invertible transformation between (U_1, U_2) and (X_1, X_2).
- The capacity of this joint ensemble is

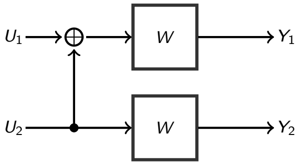
$$\begin{aligned} 2C(W) &= I(X_1X_2; Y_1Y_2) = I(U_1U_2; Y_1Y_2) \\ &= I(U_1; Y_1Y_2) + I(U_2; Y_1Y_2|U_1) \\ &= I(U_1; Y_1Y_2) + I(U_2; Y_1Y_2U_1) \end{aligned}$$



Combining Channels

- We can denote $X_1 = U_1 \oplus U_2$ and $X_2 = U_2$. There is an invertible transformation between (U_1, U_2) and (X_1, X_2).
- The capacity of this joint ensemble is

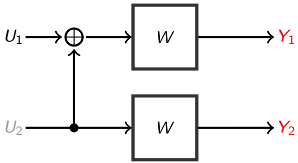
$$\begin{aligned} 2C(W) &= I(X_1X_2; Y_1Y_2) = I(U_1U_2; Y_1Y_2) \\ &= I(U_1; Y_1Y_2) + I(U_2; Y_1Y_2|U_1) \\ &= I(U_1; Y_1Y_2) + I(U_2; Y_1Y_2U_1) \\ &= C(W^-) + C(W^+) \end{aligned}$$



Combining Channels

- We can denote $X_1 = U_1 \oplus U_2$ and $X_2 = U_2$. There is an invertible transformation between (U_1, U_2) and (X_1, X_2).
- The capacity of this joint ensemble is

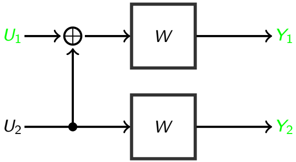
$$
\begin{aligned}
2C(W) &= I(X_1 X_2; Y_1 Y_2) = I(U_1 U_2; Y_1 Y_2) \\
&= I(U_1; Y_1 Y_2) + I(U_2; Y_1 Y_2 | U_1) \\
&= I(U_1; Y_1 Y_2) + I(U_2; Y_1 Y_2 U_1) \\
&= C(W^-) + C(W^+)
\end{aligned}
$$



Combining Channels

- We can denote $X_1 = U_1 \oplus U_2$ and $X_2 = U_2$. There is an invertible transformation between (U_1, U_2) and (X_1, X_2).
- The capacity of this joint ensemble is

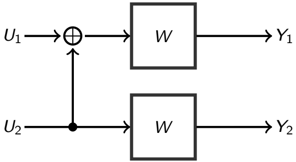
$$
\begin{aligned}
2C(W) &= I(X_1 X_2; Y_1 Y_2) = I(U_1 U_2; Y_1 Y_2) \\
&= I(U_1; Y_1 Y_2) + I(U_2; Y_1 Y_2 | U_1) \\
&= I(U_1; Y_1 Y_2) + I(U_2; Y_1 Y_2 U_1) \\
&= C(W^-) + C(W^+)
\end{aligned}
$$



Combining Channels

- We can denote $X_1 = U_1 \oplus U_2$ and $X_2 = U_2$. There is an invertible transformation between (U_1, U_2) and (X_1, X_2).
- The capacity of this joint ensemble is

$$\begin{aligned}2C(W) &= I(X_1X_2; Y_1Y_2) = I(U_1U_2; Y_1Y_2) \\ &= I(U_1; Y_1Y_2) + I(U_2; Y_1Y_2|U_1) \\ &= I(U_1; Y_1Y_2) + I(U_2; Y_1Y_2U_1) \\ &= C(W^-) + C(W^+)\end{aligned}$$



- We now have two different "channels", W^- and W^+. Total channel capacity is conserved, but distributed unevenly. One is better than the original channel, and the other worse. This is at the heart of polarization.

New Channel Capacities: BEC Case

- Channel W^- has input U_1 and output

$$(Y_1, Y_2) = \begin{cases} (X_1, X_2) & \text{w.p. } (1-p)^2 \\ (?, X_2) & \text{w.p. } p(1-p) \\ (X_1, ?) & \text{w.p. } (1-p)p \\ (?, ?) & \text{w.p. } p^2 \end{cases}$$

New Channel Capacities: BEC Case

- Channel W^- has input U_1 and output

$$(Y_1, Y_2) = \begin{cases} (X_1, X_2) & \text{w.p. } (1-p)^2 \\ (?, X_2) & \text{w.p. } p(1-p) \\ (X_1, ?) & \text{w.p. } (1-p)p \\ (?, ?) & \text{w.p. } p^2 \end{cases}$$

- First case is good, other three can be treated as erasure. Therefore W^- is BEC(p^-), where $p^- = 2p - p^2$.

New Channel Capacities: BEC Case

- Channel W^- has input U_1 and output

$$(Y_1, Y_2) = \begin{cases} (X_1, X_2) & \text{w.p. } (1-p)^2 \\ (?, X_2) & \text{w.p. } p(1-p) \\ (X_1, ?) & \text{w.p. } (1-p)p \\ (?, ?) & \text{w.p. } p^2 \end{cases}$$

- First case is good, other three can be treated as erasure. Therefore W^- is $\mathrm{BEC}(p^-)$, where $p^- = 2p - p^2$.

- Channel W^+ has input U_2 and output

$$(Y_1, Y_2, U_1) = \begin{cases} (X_1, X_2, U_1) & \text{w.p. } (1-p)^2 \\ (?, X_2, U_1) & \text{w.p. } p(1-p) \\ (X_1, ?, U_1) & \text{w.p. } (1-p)p \\ (?, ?, U_1) & \text{w.p. } p^2 \end{cases}$$

New Channel Capacities: BEC Case

- Channel W^- has input U_1 and output

$$(Y_1, Y_2) = \begin{cases} (X_1, X_2) & \text{w.p. } (1-p)^2 \\ (?, X_2) & \text{w.p. } p(1-p) \\ (X_1, ?) & \text{w.p. } (1-p)p \\ (?, ?) & \text{w.p. } p^2 \end{cases}$$

- First case is good, other three can be treated as erasure. Therefore W^- is BEC(p^-), where $p^- = 2p - p^2$.

- Channel W^+ has input U_2 and output

$$(Y_1, Y_2, U_1) = \begin{cases} (X_1, X_2, U_1) & \text{w.p. } (1-p)^2 \\ (?, X_2, U_1) & \text{w.p. } p(1-p) \\ (X_1, ?, U_1) & \text{w.p. } (1-p)p \\ (?, ?, U_1) & \text{w.p. } p^2 \end{cases}$$

- Last case can be treated as erasure, other three are good. Therefore W^+ is BEC(p^+), where $p^+ = p^2$.

New Channel Capacities: BEC Case

- Channel W^- has input U_1 and output

$$(Y_1, Y_2) = \begin{cases} (X_1, X_2) & \text{w.p. } (1-p)^2 \\ (?, X_2) & \text{w.p. } p(1-p) \\ (X_1, ?) & \text{w.p. } (1-p)p \\ (?, ?) & \text{w.p. } p^2 \end{cases}$$

- First case is good, other three can be treated as erasure. Therefore W^- is BEC(p^-), where $p^- = 2p - p^2$.

- Channel W^+ has input U_2 and output

$$(Y_1, Y_2, U_1) = \begin{cases} (X_1, X_2, U_1) & \text{w.p. } (1-p)^2 \\ (?, X_2, U_1) & \text{w.p. } p(1-p) \\ (X_1, ?, U_1) & \text{w.p. } (1-p)p \\ (?, ?, U_1) & \text{w.p. } p^2 \end{cases}$$

- Last case can be treated as erasure, other three are good. Therefore W^+ is BEC(p^+), where $p^+ = p^2$.

Note that $p^- \geq p^+$ for all $p \in [0, 1]$. Therefore, we have the relation [2]

$$C(W^-) \leq C(W) \leq C(W^+)$$

[2] In general, $C(W^+) = 2C(W) - C^2(W)$ and $C(W^-) = C^2(W)$. Refer to [1] for details.

Digression: Real Or Not?

- Channel $W^- : U_1 \to (Y_1, Y_2)$. Valid channel.

Digression: Real Or Not?

- Channel $W^- : U_1 \to (Y_1, Y_2)$. Valid channel.
- Channel $W^+ : U_2 \to (Y_1, Y_2, U_1)$.

Digression: Real Or Not?

- Channel $W^{-} : U_1 \rightarrow (Y_1, Y_2)$. Valid channel.
- Channel $W^{+} : U_2 \rightarrow (Y_1, Y_2, U_1)$. Not available to receiver.

Digression: Real Or Not?

- Channel $W^- : U_1 \to (Y_1, Y_2)$. Valid channel.
- Channel $W^+ : U_2 \to (Y_1, Y_2, U_1)$. Not available to receiver.

Theoretical Receiver **Practical Receiver**

Digression: Real Or Not?

- Channel $W^- : U_1 \to (Y_1, Y_2)$. Valid channel.
- Channel $W^+ : U_2 \to (Y_1, Y_2, U_1)$. Not available to receiver.

Theoretical Receiver

$$\hat{U}_1 = \text{Decode}(Y_1, Y_2)$$

$$\hat{U}_2 = \text{Decode}(Y_1, Y_2, U_1)$$

Practical Receiver

$$\hat{U}_1 = \text{Decode}(Y_1, Y_2)$$

$$\hat{U}_2 = \text{Decode}(Y_1, Y_2, \hat{U}_1)$$

Digression: Real Or Not?

- Channel $W^{-} : U_1 \to (Y_1, Y_2)$. Valid channel.
- Channel $W^{+} : U_2 \to (Y_1, Y_2, U_1)$. Not available to receiver.

Theoretical Receiver	**Practical Receiver**
$\hat{U}_1 = \text{Decode}(Y_1, Y_2)$	$\hat{U}_1 = \text{Decode}(Y_1, Y_2)$
$\hat{U}_2 = \text{Decode}(Y_1, Y_2, U_1)$	$\hat{U}_2 = \text{Decode}(Y_1, Y_2, \hat{U}_1)$

- Observe that $(\hat{U}_1, \hat{U}_2) \neq (U_1, U_2)$ for practical receiver (also termed **block error**) only when it is also the case with the theoretical receiver.

Digression: Real Or Not?

- Channel $W^- : U_1 \to (Y_1, Y_2)$. Valid channel.
- Channel $W^+ : U_2 \to (Y_1, Y_2, U_1)$. Not available to receiver.

Theoretical Receiver

$$\hat{U}_1 = \text{Decode}(Y_1, Y_2)$$

$$\hat{U}_2 = \text{Decode}(Y_1, Y_2, U_1)$$

Practical Receiver

$$\hat{U}_1 = \text{Decode}(Y_1, Y_2)$$

$$\hat{U}_2 = \text{Decode}(Y_1, Y_2, \hat{U}_1)$$

- Observe that $(\hat{U}_1, \hat{U}_2) \neq (U_1, U_2)$ for practical receiver (also termed **block error**) only when it is also the case with the theoretical receiver.
- Therefore, our treatment for these theoretical channels holds for those implemented in practice.

General Approach

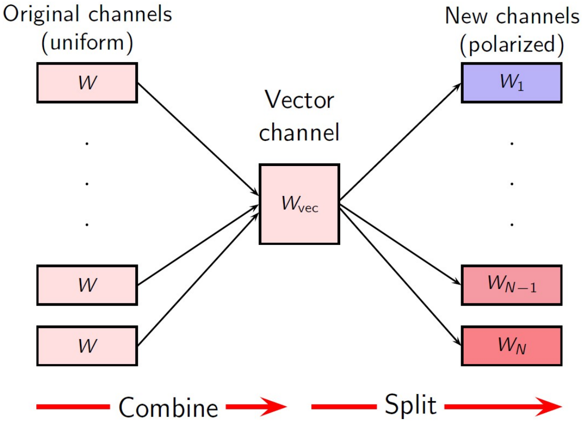
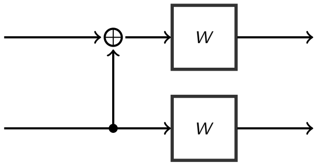
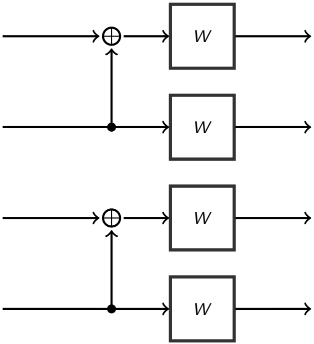


Figure: Accumulate and Redistribute Capacities (Polar Coding Tutorial, Arıkan)

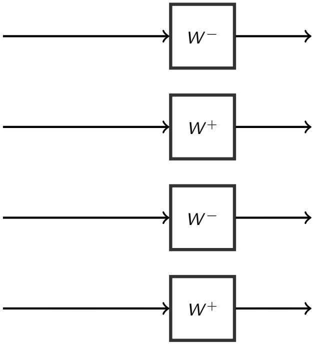
Recursively Polarize



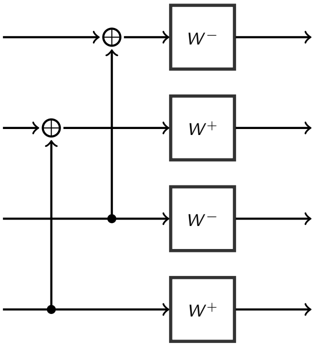
Recursively Polarize



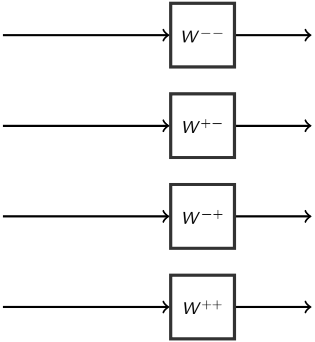
Recursively Polarize



Recursively Polarize

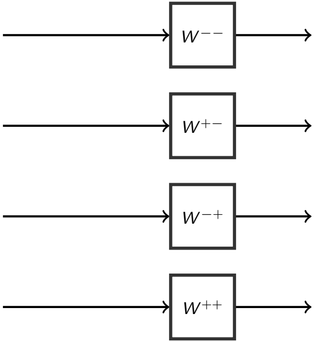


Recursively Polarize



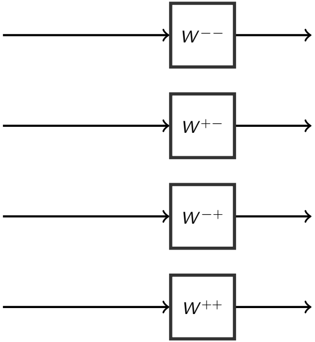
Recursively Polarize

- W^{--} is BEC(p^{--}), $p^{--} = 2p^{-} - {p^{-}}^{2}$



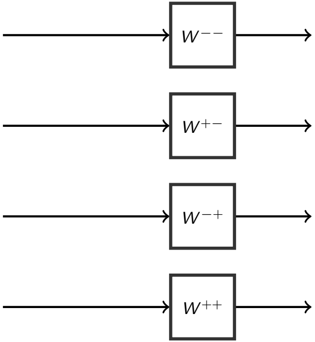
Recursively Polarize

- W^{--} is BEC(p^{--}), $p^{--} = 2p^{-} - {p^{-}}^{2}$
- W^{+-} is BEC(p^{+-}), $p^{+-} = 2p^{+} - {p^{+}}^{2}$



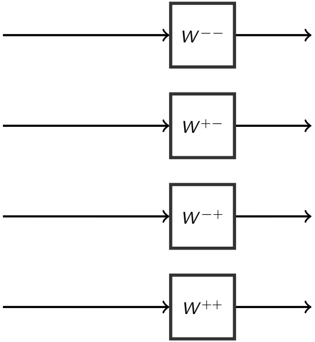
Recursively Polarize

- W^{--} is BEC(p^{--}), $p^{--} = 2p^{-} - {p^{-}}^{2}$
- W^{+-} is BEC(p^{+-}), $p^{+-} = 2p^{+} - {p^{+}}^{2}$
- W^{-+} is BEC(p^{-+}), $p^{-+} = {p^{-}}^{2}$



Recursively Polarize

- W^{--} is BEC(p^{--}), $p^{--} = 2p^{-} - {p^{-}}^2$
- W^{+-} is BEC(p^{+-}), $p^{+-} = 2p^{+} - {p^{+}}^2$
- W^{-+} is BEC(p^{-+}), $p^{-+} = {p^{-}}^2$
- W^{++} is BEC(p^{++}), $p^{++} = {p^{+}}^2$



Recursively Polarize

- W^{--} is BEC(p^{--}), $p^{--} = 2p^{-} - {p^{-}}^2$
- W^{+-} is BEC(p^{+-}), $p^{+-} = 2p^{+} - {p^{+}}^2$
- W^{-+} is BEC(p^{-+}), $p^{-+} = {p^{-}}^2$
- W^{++} is BEC(p^{++}), $p^{++} = {p^{+}}^2$

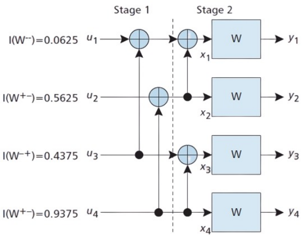


Figure: Polarized Capacities for BEC(0.5) channel [2]

Note that for $N = 2^t$ channels, we require t stages of repeated polar transforms. We can see that the capacity of W^{++} has improved significantly, while that of W^{--} has deteriorated vastly [3].

[3] Termed the **Matthew Effect:** The good become better and the bad get worse.

Recursively Polarize

- W^{--} is BEC(p^{--}), $p^{--} = 2p^{-} - {p^{-}}^2$
- W^{+-} is BEC(p^{+-}), $p^{+-} = 2p^{+} - {p^{+}}^2$
- W^{-+} is BEC(p^{-+}), $p^{-+} = {p^{-}}^2$
- W^{++} is BEC(p^{++}), $p^{++} = {p^{+}}^2$

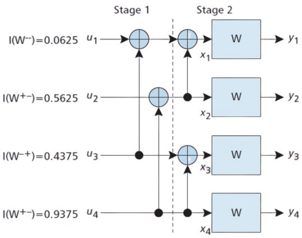


Figure: Polarized Capacities for BEC(0.5) channel [2]

Note that for $N = 2^t$ channels, we require t stages of repeated polar transforms. We can see that the capacity of W^{++} has improved significantly, while that of W^{--} has deteriorated vastly [3].
Are we approaching extremal polarization?

[3] Termed the **Matthew Effect:** The good become better and the bad get worse.

Visualizing Polarization

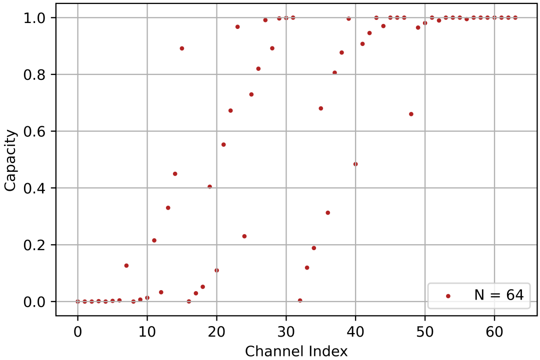


Figure: Distribution of capacities as polarization increases for BEC(0.4)

Visualizing Polarization

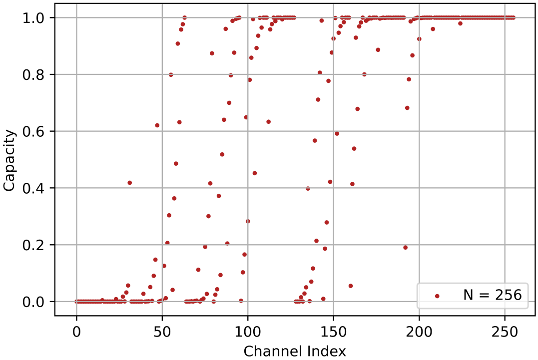


Figure: Distribution of capacities as polarization increases for BEC(0.4)

Visualizing Polarization

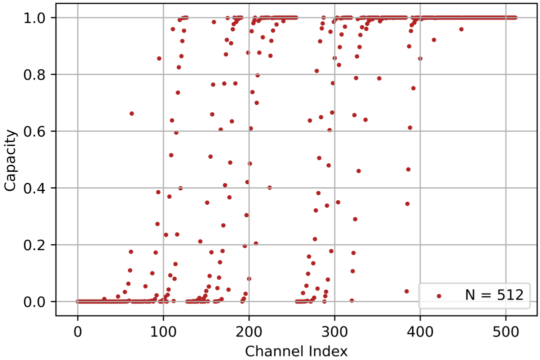


Figure: Distribution of capacities as polarization increases for BEC(0.4)

Visualizing Polarization

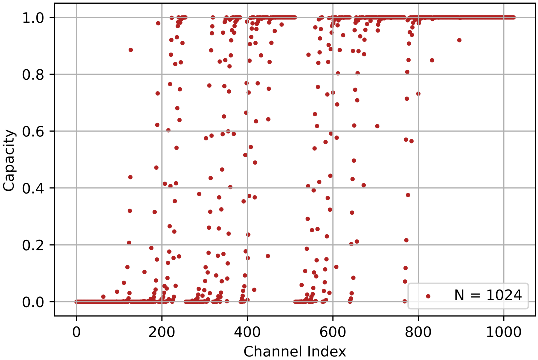


Figure: Distribution of capacities as polarization increases for BEC(0.4)

Visualizing Polarization

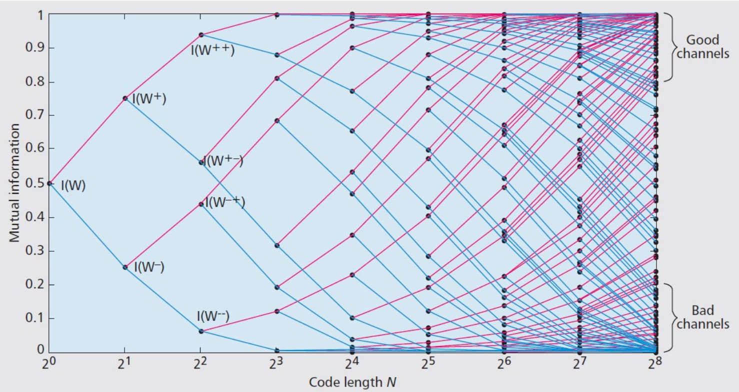


Figure: Evolution of Polarization Martingale [2]

Polarization Theorem

Theorem

As the number of channels $N = 2^t$ increases, the channel capacities $\{C(W_i)\}$ polarize. For any $\delta \in (0,1)$,

$$\frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\{C(W_i) > 1-\delta\} \longrightarrow C(W)$$

and

$$\frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\{C(W_i) < \delta\} \longrightarrow 1-C(W)$$

Polarization Theorem

Theorem

As the number of channels $N = 2^t$ increases, the channel capacities $\{C(W_i)\}$ polarize. For any $\delta \in (0,1)$,

$$\frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\{C(W_i) > 1-\delta\} \longrightarrow C(W)$$

and

$$\frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\{C(W_i) < \delta\} \longrightarrow 1 - C(W)$$

- The theorem tells us that, upon repeatedly applying polarization, the fraction of δ-good channels converges to $C(W)$, and the fraction of δ-bad channels converges to $1 - C(W)$.

Polarization Theorem

Theorem

As the number of channels $N = 2^t$ increases, the channel capacities $\{C(W_i)\}$ polarize. For any $\delta \in (0,1)$,

$$\frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\{C(W_i) > 1-\delta\} \longrightarrow C(W)$$

and

$$\frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\{C(W_i) < \delta\} \longrightarrow 1 - C(W)$$

- The theorem tells us that, upon repeatedly applying polarization, the fraction of δ-good channels converges to $C(W)$, and the fraction of δ-bad channels converges to $1 - C(W)$.
- As a corollary, we can see that the fraction of δ-"mediocre" channels, i.e. those with $C(W_i) \in (\delta, 1-\delta)$, converges to 0.

Polarization Theorem

Theorem

As the number of channels $N = 2^t$ increases, the channel capacities $\{C(W_i)\}$ polarize. For any $\delta \in (0, 1)$,

$$\frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\{C(W_i) > 1 - \delta\} \longrightarrow C(W)$$

and

$$\frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\{C(W_i) < \delta\} \longrightarrow 1 - C(W)$$

- The theorem tells us that, upon repeatedly applying polarization, the fraction of δ-good channels converges to $C(W)$, and the fraction of δ-bad channels converges to $1 - C(W)$.
- As a corollary, we can see that the fraction of δ-"mediocre" channels, i.e. those with $C(W_i) \in (\delta, 1 - \delta)$, converges to 0.
- This can be proven in a more general case by applying Doob's Martingale Convergence Theorem, and taking into account the fact that $\{C(W_i)\}$ is a martingale bounded in $(0, 1)$.

Proof of Theorem (for BEC) - I

- Let us quantify a particular behavior of a channel. For a BEC(p) channel W, define $U(W) = \sqrt{p(1-p)}$.

Proof of Theorem (for BEC) - I

- Let us quantify a particular behavior of a channel. For a BEC(p) channel W, define $U(W) = \sqrt{p(1-p)}$.
- Observe that $U(W^+) = U(W)\sqrt{p(1+p)}$ and $U(W^-) = U(W)\sqrt{(1-p)(2-p)}$.

Proof of Theorem (for BEC) - I

- Let us quantify a particular behavior of a channel. For a BEC(p) channel W, define $U(W) = \sqrt{p(1-p)}$.
- Observe that $U(W^+) = U(W)\sqrt{p(1+p)}$ and $U(W^-) = U(W)\sqrt{(1-p)(2-p)}$.
- Therefore

$$\frac{1}{2}(U(W^+) + U(W^-)) = \frac{1}{2}U(W)(\sqrt{p(1+p)} + \sqrt{(1-p)(2-p)}) \leq \frac{\sqrt{3}}{2}U(W)$$

.

Proof of Theorem (for BEC) - I

- Let us quantify a particular behavior of a channel. For a BEC(p) channel W, define $U(W) = \sqrt{p(1-p)}$.
- Observe that $U(W^+) = U(W)\sqrt{p(1+p)}$ and $U(W^-) = U(W)\sqrt{(1-p)(2-p)}$.
- Therefore

$$\frac{1}{2}(U(W^+) + U(W^-)) = \frac{1}{2}U(W)(\sqrt{p(1+p)} + \sqrt{(1-p)(2-p)}) \leq \frac{\sqrt{3}}{2}U(W)$$

.

- Expanding the sum

$$\frac{1}{2^t}\sum_{i=1}^{2^t} U(W_i) \leq \left(\frac{\sqrt{3}}{2}\right)^t U(W)$$

Proof of Theorem (for BEC) - I

- Let us quantify a particular behavior of a channel. For a BEC(p) channel W, define $U(W) = \sqrt{p(1-p)}$.
- Observe that $U(W^+) = U(W)\sqrt{p(1+p)}$ and $U(W^-) = U(W)\sqrt{(1-p)(2-p)}$.
- Therefore

$$\frac{1}{2}(U(W^+) + U(W^-)) = \frac{1}{2}U(W)(\sqrt{p(1+p)} + \sqrt{(1-p)(2-p)}) \leq \frac{\sqrt{3}}{2}U(W)$$

.

- Expanding the sum

$$\frac{1}{2^t}\sum_{i=1}^{2^t} U(W_i) \leq \left(\frac{\sqrt{3}}{2}\right)^t U(W)$$

- We also have that

$$\mathbb{1}\{C(W_i) \in (\delta, 1-\delta)\} \leq \frac{U(W_i)}{\sqrt{\delta(1-\delta)}}$$

Proof of Theorem (for BEC) - II

- If we define the fraction of δ-mediocre channels as $\mu_t(\delta)$, we have

$$\mu_t(\delta) := \frac{1}{2^t} \sum_{i=1}^{2^t} \mathbb{1}\{C(W_i) \in (\delta, 1-\delta)\}$$

Proof of Theorem (for BEC) - II

- If we define the fraction of δ-mediocre channels as $\mu_t(\delta)$, we have

$$\mu_t(\delta) := \frac{1}{2^t} \sum_{i=1}^{2^t} \mathbb{1}\{C(W_i) \in (\delta, 1-\delta)\}$$

- Therefore

$$\mu_t(\delta) \le \left(\frac{\sqrt{3}}{2}\right)^t \frac{U(W)}{\sqrt{\delta(1-\delta)}}$$

Proof of Theorem (for BEC) - II

- If we define the fraction of δ-mediocre channels as $\mu_t(\delta)$, we have

$$\mu_t(\delta) := \frac{1}{2^t} \sum_{i=1}^{2^t} \mathbb{1}\{C(W_i) \in (\delta, 1-\delta)\}$$

- Therefore

$$\mu_t(\delta) \leq \left(\frac{\sqrt{3}}{2}\right)^t \frac{U(W)}{\sqrt{\delta(1-\delta)}}$$

- So as the number of channels N (and hence t) increases, this fraction goes to zero.

Proof of Theorem (for BEC) - II

- If we define the fraction of δ-mediocre channels as $\mu_t(\delta)$, we have

$$\mu_t(\delta) := \frac{1}{2^t} \sum_{i=1}^{2^t} \mathbb{1}\{C(W_i) \in (\delta, 1-\delta)\}$$

- Therefore

$$\mu_t(\delta) \leq \left(\frac{\sqrt{3}}{2}\right)^t \frac{U(W)}{\sqrt{\delta(1-\delta)}}$$

- So as the number of channels N (and hence t) increases, this fraction goes to zero.
- The fraction of perfect and useless channels follows trivially from the conservation of channel capacity.

$\blacksquare$

Encoding Scheme

Suppose we wish to communicate at rate R:

Encoding Scheme

Suppose we wish to communicate at rate R:

- Create $N = 2^t$ copies of the original channel W. We can apply the polarization transformation t times to generate our N synthetic channels.

Encoding Scheme

Suppose we wish to communicate at rate R:

- Create $N = 2^t$ copies of the original channel W. We can apply the polarization transformation t times to generate our N synthetic channels.
- Select the $k = N \cdot R$ synthetic channels with the best polarized channel capacities, and set their inputs to be our information bits. Freeze the inputs of the remaining channels to some value (say 0).

Encoding Scheme

Suppose we wish to communicate at rate R:

- Create $N = 2^t$ copies of the original channel W. We can apply the polarization transformation t times to generate our N synthetic channels.
- Select the $k = N \cdot R$ synthetic channels with the best polarized channel capacities, and set their inputs to be our information bits. Freeze the inputs of the remaining channels to some value (say 0).
- Note that we can almost continuously adjust the rate of transmission by adding or deleting a polarized channel in use.

Encoding Scheme

Suppose we wish to communicate at rate R:

- Create $N = 2^t$ copies of the original channel W. We can apply the polarization transformation t times to generate our N synthetic channels.
- Select the $k = N \cdot R$ synthetic channels with the best polarized channel capacities, and set their inputs to be our information bits. Freeze the inputs of the remaining channels to some value (say 0).
- Note that we can almost continuously adjust the rate of transmission by adding or deleting a polarized channel in use.
- Furthermore, the polarization theorem guarantees that, for large enough N, the fraction of perfect channels approaches $C(W)$. In such a scenario, we can reliably select these perfect channels to be our best ones to transmit information bits over. Therefore, we can achieve rates all the way up till the channel capacity.

Encoding Scheme

Suppose we wish to communicate at rate R:

- Create $N = 2^t$ copies of the original channel W. We can apply the polarization transformation t times to generate our N synthetic channels.
- Select the $k = N \cdot R$ synthetic channels with the best polarized channel capacities, and set their inputs to be our information bits. Freeze the inputs of the remaining channels to some value (say 0).
- Note that we can almost continuously adjust the rate of transmission by adding or deleting a polarized channel in use.
- Furthermore, the polarization theorem guarantees that, for large enough N, the fraction of perfect channels approaches $C(W)$. In such a scenario, we can reliably select these perfect channels to be our best ones to transmit information bits over. Therefore, we can achieve rates all the way up till the channel capacity.

Remark

Unlike the goal of traditional code construction to maximize the minimum distance between codewords, polar coding aims to reduce the probability of error along information-bearing channels.

Encoding Example

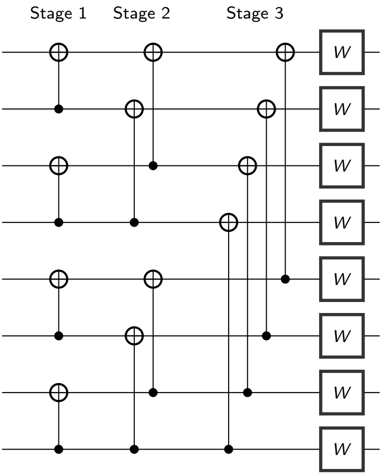


Figure: 3-stage polar encoder for BEC(0.4)

Encoding Example

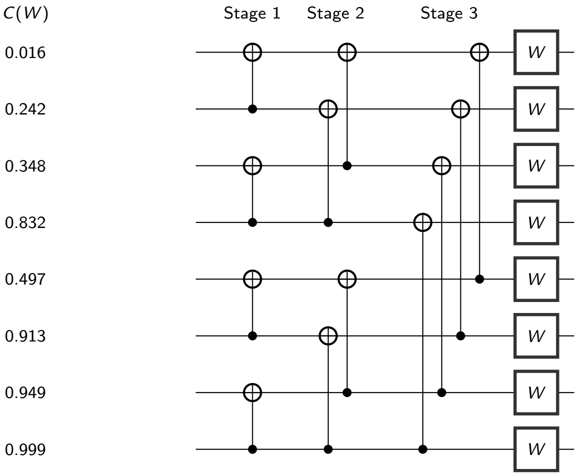


Figure: 3-stage polar encoder for BEC(0.4)

Encoding Example

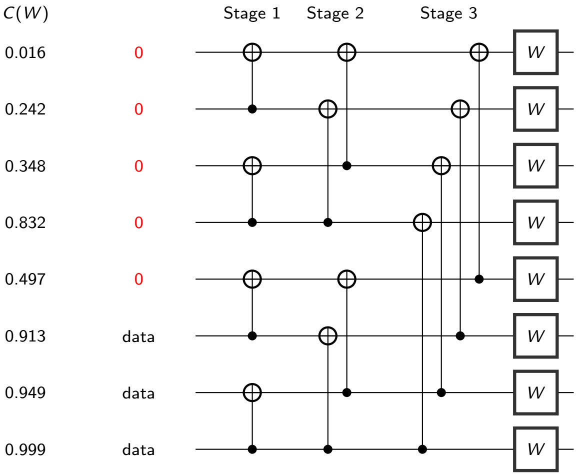


Figure: 3-stage polar encoder for BEC(0.4)

Encoding Example

- Suppose we maintain a low tolerance of error, and we decide to use only the best three synthesized channels for transmitting data. Therefore, we have

$$\text{Rate} = \frac{3}{8} = 0.375$$

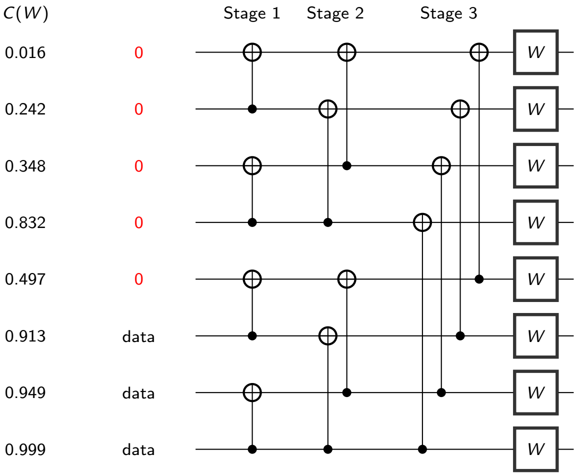


Figure: 3-stage polar encoder for BEC(0.4)

Encoding Example

- Suppose we maintain a low tolerance of error, and we decide to use only the best three synthesized channels for transmitting data. Therefore, we have

$$\text{Rate} = \frac{3}{8} = 0.375$$

- Our control over the rate can be seen more clearly if we decide to use the best 4 channels. In this case

$$\text{Rate} = \frac{4}{8} = 0.5$$

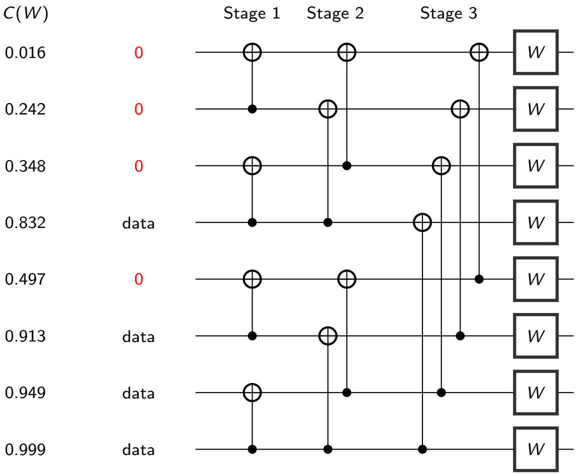


Figure: 3-stage polar encoder for BEC(0.4)

Encoding Complexity

For a code of length $N = 2^t$,

- As noted, we will have t polarization stages to synthesize our channels.

Encoding Complexity

For a code of length $N = 2^t$,

- As noted, we will have t polarization stages to synthesize our channels.
- Each polarization stage has $N/2$ **butterfly units** so as to cover all codeword indices. Each butterfly unit can carry out computation in $\mathcal{O}(1)$ time.

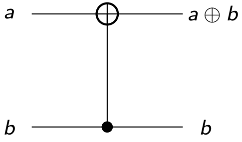


Figure: Butterfly Unit

Encoding Complexity

For a code of length $N = 2^t$,

- As noted, we will have t polarization stages to synthesize our channels.
- Each polarization stage has $N/2$ **butterfly units** so as to cover all codeword indices. Each butterfly unit can carry out computation in $\mathcal{O}(1)$ time.
- Therefore, encoding complexity is $\mathcal{O}(t \cdot N)$, or

$$\mathcal{O}(N \log N)$$

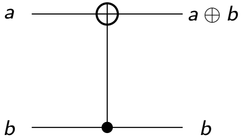
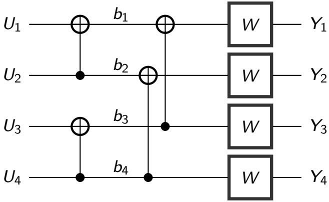


Figure: Butterfly Unit

Successive Cancellation Decoding

Looking at Stage 2

- $W^{-} : b_1 \rightarrow Y_1 Y_3$
- $W^{-} : b_2 \rightarrow Y_2 Y_4$
- $W^{+} : b_3 \rightarrow Y_1 Y_3 \hat{b_1}$
- $W^{+} : b_4 \rightarrow Y_2 Y_4 \hat{b_2}$



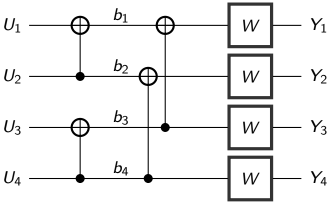
Successive Cancellation Decoding

Looking at Stage 2

- $W^{-} : b_1 \rightarrow Y_1 Y_3$
- $W^{-} : b_2 \rightarrow Y_2 Y_4$
- $W^{+} : b_3 \rightarrow Y_1 Y_3 \hat{b_1}$
- $W^{+} : b_4 \rightarrow Y_2 Y_4 \hat{b_2}$

Looking at Stage 1

- $W^{--} : U_1 \rightarrow b_1 b_2$
- $W^{-+} : U_2 \rightarrow b_1 b_2 \hat{U}_1$
- $W^{+-} : U_3 \rightarrow b_3 b_4$
- $W^{++} : U_4 \rightarrow b_3 b_4 \hat{U}_3$



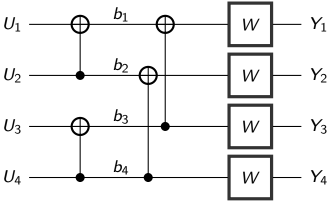
Successive Cancellation Decoding

Looking at Stage 2

- $W^- : b_1 \to Y_1 Y_3$
- $W^- : b_2 \to Y_2 Y_4$
- $W^+ : b_3 \to Y_1 Y_3 \hat{b}_1$
- $W^+ : b_4 \to Y_2 Y_4 \hat{b}_2$

Looking at Stage 1

- $W^{--} : U_1 \to b_1 b_2$
- $W^{-+} : U_2 \to b_1 b_2 \hat{U}_1$
- $W^{+-} : U_3 \to b_3 b_4$
- $W^{++} : U_4 \to b_3 b_4 \hat{U}_3$



Therefore, our channels can be represented as

- $W^{--} : U_1 \to Y_1 Y_2 Y_3 Y_4$
- $W^{-+} : U_2 \to Y_1 Y_2 Y_3 Y_4 \hat{U}_1$
- $W^{+-} : U_3 \to Y_1 Y_2 Y_3 Y_4 \hat{b}_1 \hat{b}_2$
- $W^{++} : U_4 \to Y_1 Y_2 Y_3 Y_4 \hat{b}_1 \hat{b}_2 \hat{U}_3$

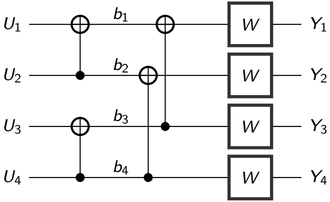
Successive Cancellation Decoding

Looking at Stage 2

- $W^{-} : b_1 \rightarrow Y_1 Y_3$
- $W^{-} : b_2 \rightarrow Y_2 Y_4$
- $W^{+} : b_3 \rightarrow Y_1 Y_3 \hat{b_1}$
- $W^{+} : b_4 \rightarrow Y_2 Y_4 \hat{b_2}$

Looking at Stage 1

- $W^{--} : U_1 \rightarrow b_1 b_2$
- $W^{-+} : U_2 \rightarrow b_1 b_2 \hat{U}_1$
- $W^{+-} : U_3 \rightarrow b_3 b_4$
- $W^{++} : U_4 \rightarrow b_3 b_4 \hat{U}_3$



Therefore, our channels can be represented as

- $W^{--} : U_1 \rightarrow Y_1 Y_2 Y_3 Y_4$
- $W^{-+} : U_2 \rightarrow Y_1 Y_2 Y_3 Y_4 \hat{U}_1$
- $W^{+-} : U_3 \rightarrow Y_1 Y_2 Y_3 Y_4 \hat{U}_1 \hat{U}_2$
- $W^{++} : U_4 \rightarrow Y_1 Y_2 Y_3 Y_4 \hat{U}_1 \hat{U}_2 \hat{U}_3$

Algorithm and Complexity

- Decoding can be done using maximum likelihood decision making.

$$L^{--} = \frac{\mathbb{P}\{Y_1 Y_2 Y_3 Y_4 | U_1 = 0\}}{\mathbb{P}\{Y_1 Y_2 Y_3 Y_4 | U_1 = 1\}} \qquad L^{-+} = \frac{\mathbb{P}\{Y_1 Y_2 Y_3 Y_4 \hat{U}_1 | U_2 = 0\}}{\mathbb{P}\{Y_1 Y_2 Y_3 Y_4 \hat{U}_1 | U_2 = 1\}}$$

$$\hat{U}_1 = \begin{cases} 0 & \text{if } U_1 \text{ is frozen} \\ 0 & \text{if } L^{--} > 1 \\ 1 & \text{otherwise} \end{cases} \qquad \hat{U}_2 = \begin{cases} 0 & \text{if } U_2 \text{ is frozen} \\ 0 & \text{if } L^{-+} > 1 \\ 1 & \text{otherwise} \end{cases}$$

Decisions for $\hat{U}_3$ and $\hat{U}_4$ follow similarly.

Algorithm and Complexity

- Decoding can be done using maximum likelihood decision making.

$$L^{--} = \frac{\mathbb{P}\{Y_1 Y_2 Y_3 Y_4 | U_1 = 0\}}{\mathbb{P}\{Y_1 Y_2 Y_3 Y_4 | U_1 = 1\}} \qquad L^{-+} = \frac{\mathbb{P}\{Y_1 Y_2 Y_3 Y_4 \hat{U}_1 | U_2 = 0\}}{\mathbb{P}\{Y_1 Y_2 Y_3 Y_4 \hat{U}_1 | U_2 = 1\}}$$

$$\hat{U}_1 = \begin{cases} 0 & \text{if } U_1 \text{ is frozen} \\ 0 & \text{if } L^{--} > 1 \\ 1 & \text{otherwise} \end{cases} \qquad \hat{U}_2 = \begin{cases} 0 & \text{if } U_2 \text{ is frozen} \\ 0 & \text{if } L^{-+} > 1 \\ 1 & \text{otherwise} \end{cases}$$

  Decisions for $\hat{U}_3$ and $\hat{U}_4$ follow similarly.

- We have $t = \log N$ stages, and at each stage we make N estimations. Therefore, our decoding complexity is

$$\mathcal{O}(N \log N)$$

References

E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3051–3073, 2009.

K. Niu, K. Chen, J. Lin, and Q. T. Zhang, "Polar codes: Primary concepts and practical decoding algorithms," *IEEE Communications Magazine*, vol. 52, no. 7, pp. 192–203, 2014.

Relevant Resources:

- The Flesh of Polar Codes, ISIT 2017
- Polar Coding Tutorial, Arıkan

Algebraic Formulation

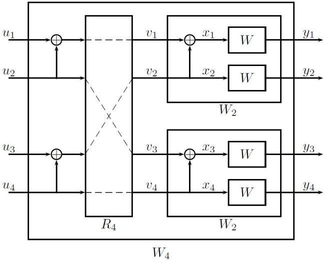


Figure: Symmetric block diagram for 2-stage polarizer [1]

Algebraic Formulation

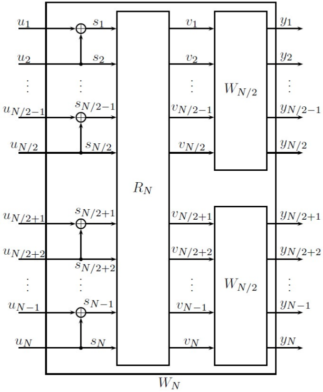


Figure: Symmetric block diagram for general polarizer [1]

Algebraic Formulation

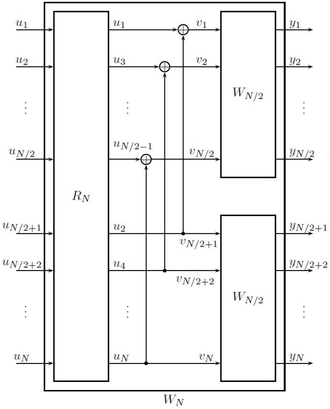


Figure: Alternate block diagram for general polarizer [1]

Algebraic Formulation

- Define

$$F_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

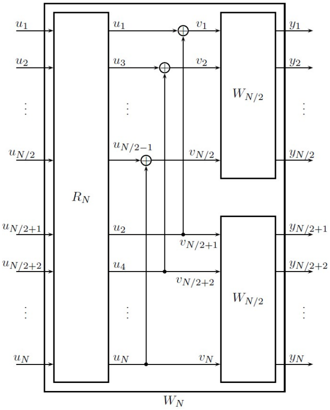


Figure: Alternate block diagram for general polarizer [1]

Algebraic Formulation

- Define

$$F_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

- Then, for $N = 2^t$,

$$F_{2N} = \begin{bmatrix} F_N & 0 \\ F_N & F_N \end{bmatrix}$$

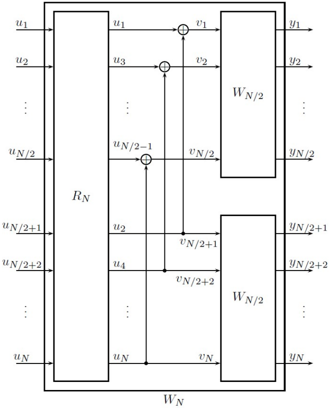


Figure: Alternate block diagram for general polarizer [1]

Algebraic Formulation

- Define

$$F_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

- Then, for $N = 2^t$,

$$F_{2N} = \begin{bmatrix} F_N & 0 \\ F_N & F_N \end{bmatrix}$$

- Therefore, the generator matrix for a $N = 2^t$ length polar code is

$$G_N = B_N F_N$$

where B_N is a bit-reversal permutation matrix.

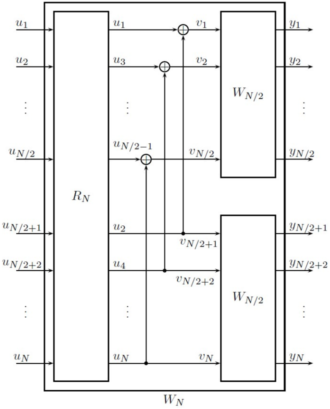


Figure: Alternate block diagram for general polarizer [1]