EE376A
Information Theory

Lecture 9: Polar Codes

Mert Pilanci

Stanford University

February 5, 2019

Outline

- Channel coding and capacity
- Polar code construction
- Decoding
- Theoretical analysis
- Extensions

Channel coding

- **Entropy**

$$H(U) = \mathbb{E}[\log \frac{1}{p(U)}] = -\sum_{u} p(u) \log p(u)$$

Channel coding

- **Entropy**

$$H(U) = \mathbb{E}[\log \frac{1}{p(U)}] = -\sum_u p(u) \log p(u)$$

- **Conditional Entropy**

$$H(X|Y) = \mathbb{E}[\log \frac{1}{p(X|Y)}] = \sum_y p(y) H(X|Y = y)$$

Channel coding

- **Entropy**

$$H(U) = \mathbb{E}[\log \frac{1}{p(U)}] = -\sum_{u} p(u) \log p(u)$$

- **Conditional Entropy**

$$H(X|Y) = \mathbb{E}[\log \frac{1}{p(X|Y)}] = \sum_{y} p(y) H(X|Y=y)$$

- **Mutual Information**

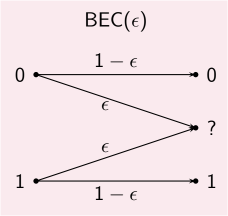
$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \\ &= H(X) + H(Y) - H(X,Y) \end{aligned}$$

Channel Capacity

- Channel capacity C is the maximal rate of reliable communication over memoryless channel characterized by $P(Y|X)$
- Theorem:

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

Capacity of the binary erasure channel (BEC)



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

Picking $X \sim Ber(\frac{1}{2})$, we have $H(X) = 1$. Thus, the capacity of BEC is $C = 1 - \epsilon$

Channel Coding

$$J \sim \text{uniform} \in \{1, 2, ..., M\} \to \boxed{\text{encoder}} \xrightarrow{X^n} \boxed{\text{memoryless channel } P_{Y|X}} \xrightarrow{Y^n} \boxed{\text{decoder}} \to \hat{J}$$

rate $\frac{\log M}{n}$ bits/channel use

probability of error $P_e = P(\hat{J} \neq J)$

Channel Coding

$$J \sim \text{uniform} \in \{1, 2, ..., M\} \rightarrow \boxed{\text{encoder}} \xrightarrow{X^n} \boxed{\text{memoryless channel } P_{Y|X}} \xrightarrow{Y^n} \boxed{\text{decoder}} \rightarrow \hat{J}$$

rate $\frac{\log M}{n}$ bits/channel use

probability of error $P_e = P(\hat{J} \neq J)$

- If $R < \max_{P_x} I(X;Y)$, then rate R is achievable, i.e., there exists schemes with rate $\geq R$ and $P_e \rightarrow 0$

Channel Coding

$$J \sim \text{uniform} \in \{1, 2, ..., M\} \rightarrow \boxed{\text{encoder}} \xrightarrow{X^n} \boxed{\text{memoryless channel } P_{Y|X}} \xrightarrow{Y^n} \boxed{\text{decoder}} \rightarrow \hat{J}$$

rate $\frac{\log M}{n}$ bits/channel use

probability of error $P_e = P(\hat{J} \neq J)$

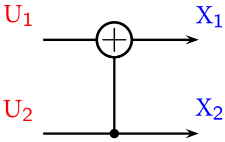
- If $R < \max_{P_x} I(X;Y)$, then rate R is achievable, i.e., there exists schemes with rate $\geq R$ and $P_e \rightarrow 0$
- If $R > \max_{P_x} I(X;Y)$, then R is not achievable.

Main result: maximum rate of reliable communication $C = \max_{P_X} I(X;Y)$

Today: Polar Codes

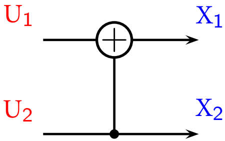
- Invented by Erdal Arikan in 2009
- First code with an explicit construction to provably achieve the channel capacity
- Nice structure with efficient encoding/decoding operations
- We will assume that the channel is symmetric, i.e., uniform input distribution achieves capacity

Basic 2×2 transformation



$U_1, U_2, X_1, X_2 \in \{0, 1\}$ binary variables (in GF(2))

Basic 2×2 transformation

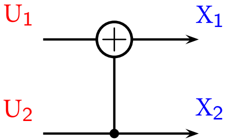


$U_1, U_2, X_1, X_2 \in \{0, 1\}$ binary variables (in GF(2))

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \mod 2$$

or equivalently $X_1 = U_1 \oplus U_2$ and $X_2 = U_2$

Properties of G_2

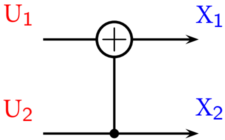


$$U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \qquad X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

Define $G_2 := \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ then we have $X = G_2 U$

$$G_2^2 := G_2 G_2$$

Properties of G_2



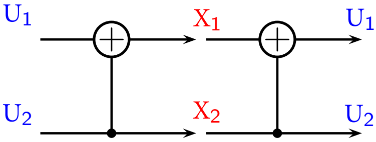
$$U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \qquad X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

Define $\quad G_2 := \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ then we have $X = G_2 U$

$$G_2^2 := G_2 G_2$$

$$G_2 G_2 U = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix}$$

Properties of G_2^2



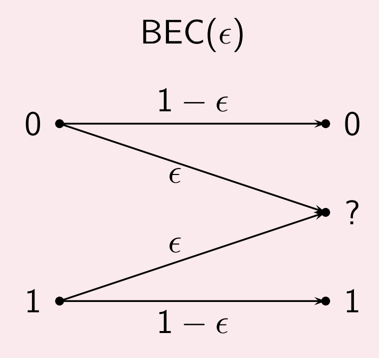
Define $\quad G_2 := \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ then we have $X = G_2 U$

$$G_2^2 := G_2 G_2$$

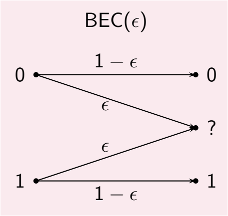
$$\begin{aligned} G_2 G_2 U &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} U_1 \oplus U_2 \\ U_2 \end{bmatrix} \end{aligned}$$

$=$

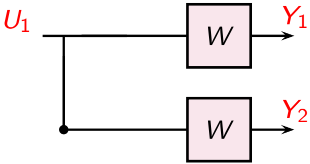
Erasure channel



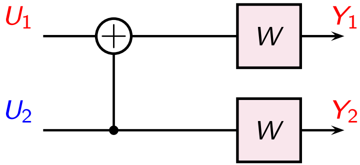
Naively combining erasure channels



- Repetition coding



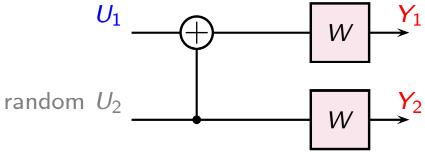
Combining two erasure channels



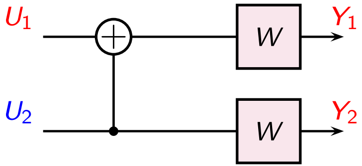
Invertible transformation does not alter capacity:
$I(U;Y) = I(X;Y)$

Sequential decoding

First bit-channel $W_1 : U_1 \to (Y_1, Y_2)$



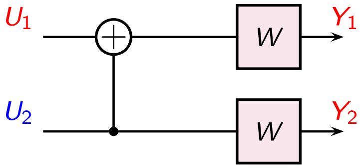
Second bit-channel $W_2 : U_2 \to (Y_1, Y_2, U_1)$



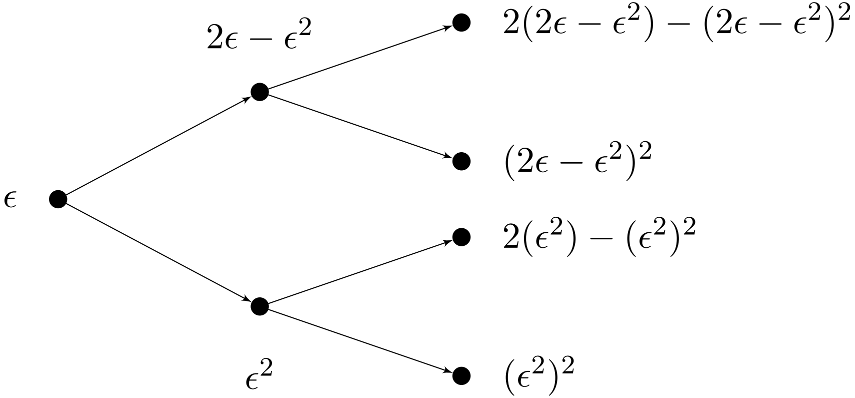
Capacity is conserved

$$C(W_1) + C(W_2) = C(W) + C(W) = 2C(W)$$

$$C(W_1) \leq C(W) \leq C(W_2)$$



Polarization process



A familiar update rule...

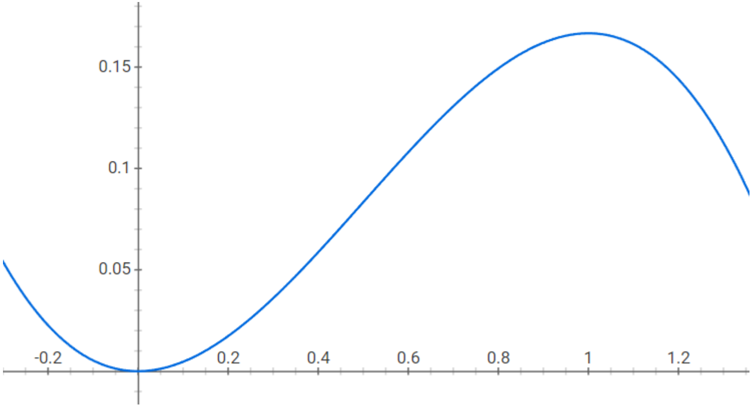
Let e_t be i.i.d. uniform ± 1 for $t = 1, 2...$

$$w_{t+1} = w_t + e_t w_t (1 - w_t)$$

A familiar update rule

Let e_t be i.i.d. uniform ± 1 for $t = 1, 2...$

$$w_{t+1} = w_t + e_t w_t (1 - w_t)$$



Martingales

Let e_t be i.i.d. uniform ± 1 for $t = 1, 2...$

$$w_{t+1} = w_t + e_t w_t (1 - w_t)$$

$$\mathbb{E}[w_{t+1} | w_t] = w_t$$

Martingales

Let e_t be i.i.d. uniform ± 1 for $t = 1, 2...$

$$w_{t+1} = w_t + e_t w_t (1 - w_t)$$

$$\mathbb{E}[w_{t+1} | w_t] = w_t$$

- Doob's Martingale convergence theorem

  (informal) Bounded Martingale processes converge to a limiting random variable w_∞ such that $\mathbb{E}[|w_t - w_\infty|] \to 0$.

Non-convergent paths

- Down - Up - Down - Up ....
  $\epsilon \searrow \epsilon^2 \nearrow 2\epsilon^2 - \epsilon^4 =? \epsilon$

Non-convergent paths

- Down - Up - Down - Up ....

  $\epsilon \searrow \epsilon^2 \nearrow 2\epsilon^2 - \epsilon^4 = \epsilon$ if $\epsilon = \frac{\sqrt{5}}{2} - \frac{1}{2} = \frac{1}{\phi} \approx 0.61803398875$

Non-convergent paths

- Down - Up - Down - Up ....

$\epsilon \searrow \epsilon^2 \nearrow 2\epsilon^2 - \epsilon^4 = \epsilon$ if $\epsilon = \frac{\sqrt{5}}{2} - \frac{1}{2} = \frac{1}{\phi} \approx 0.61803398875$

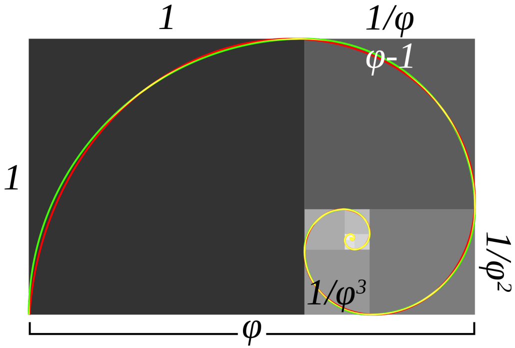
$$\text{Golden ratio :} \quad \phi := \frac{1+\sqrt{5}}{2} \approx 1.61803398875$$

Non-convergent paths

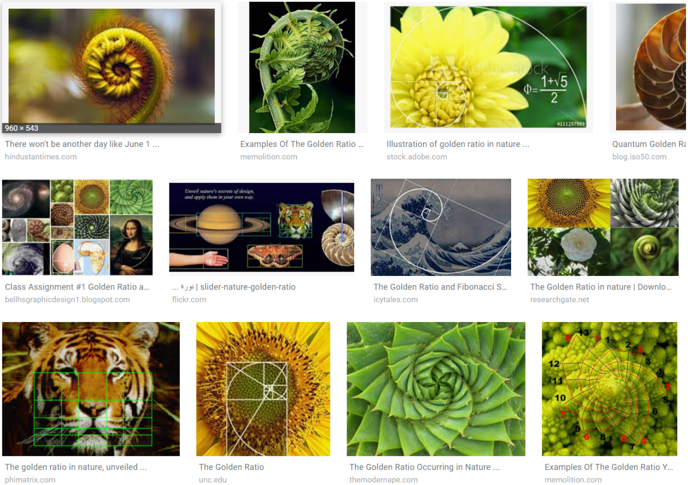
- Down - Up - Down - Up ....

$\epsilon \searrow \epsilon^2 \nearrow 2\epsilon^2 - \epsilon^4 = \epsilon$ if $\epsilon = \frac{\sqrt{5}}{2} - \frac{1}{2} = \frac{1}{\phi} \approx 0.61803398875$

$$\text{Golden ratio :} \quad \phi := \frac{1+\sqrt{5}}{2} \approx 1.61803398875$$



Google images: golden ratio in nature



Polarization theorem

Theorem

The bit-channel capacities $\{C(W_i)\}$ *polarize: for any* $\delta \in (0,1)$, *as the construction size* N *grows*

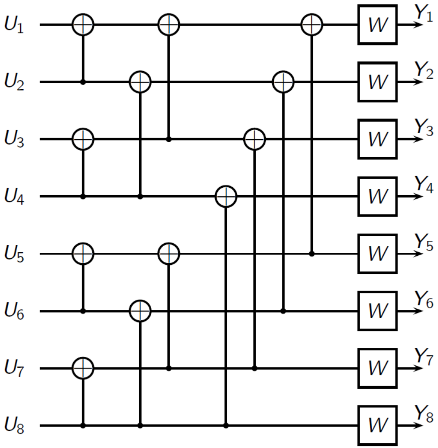
$$\left[\frac{\text{no. channels with } C(W_i) > 1-\delta}{N}\right] \longrightarrow C(W)$$

and

$$\left[\frac{\text{no. channels with } C(W_i) < \delta}{N}\right] \longrightarrow 1 - C(W)$$

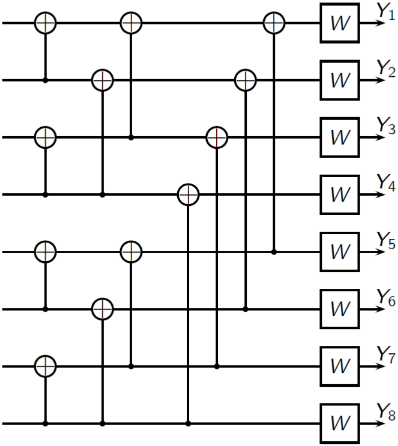
Freezing noisy channels

$I(W_i)$	Rank
0.0039	8
0.1211	7
0.1914	6
0.6836	4
0.3164	5
0.8086	3
0.8789	2
0.9961	1



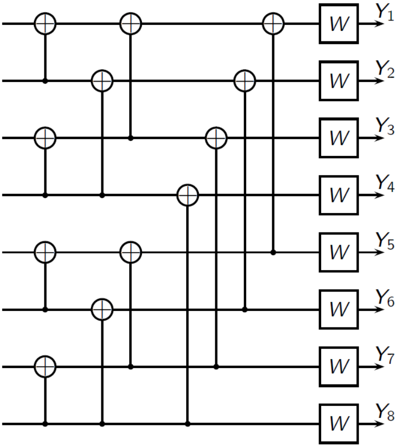
Freezing noisy channels

$I(W_i)$	Rank		
0.0039	8	frozen	U_1
0.1211	7	frozen	U_2
0.1914	6	frozen	U_3
0.6836	4	data	U_4
0.3164	5	frozen	U_5
0.8086	3	data	U_6
0.8789	2	data	U_7
0.9961	1	data	U_8

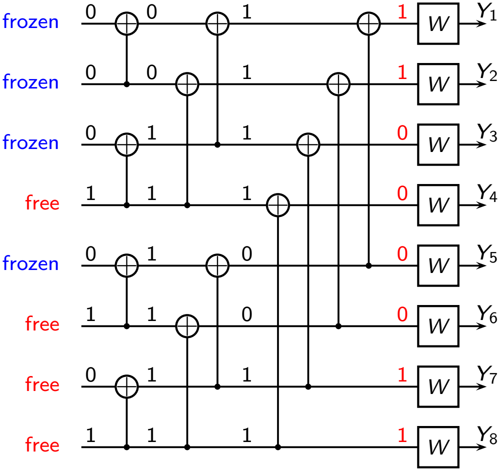


Freezing noisy channels

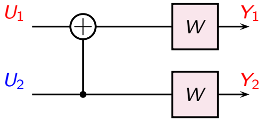
$I(W_i)$	Rank		
0.0039	8	frozen	0
0.1211	7	frozen	0
0.1914	6	frozen	0
0.6836	4	data	U_4
0.3164	5	frozen	0
0.8086	3	data	U_6
0.8789	2	data	U_7
0.9961	1	data	U_8



Encoding


frozen 0 0 1 1 W Y_1
frozen 0 0 1 1 W Y_2
frozen 0 1 1 0 W Y_3
free 1 1 1 0 W Y_4
frozen 0 1 0 0 W Y_5
free 1 1 0 0 W Y_6
free 0 1 1 1 W Y_7
free 1 1 1 1 W Y_8

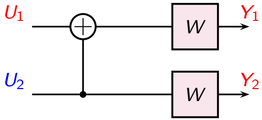
Polarization of general channels



$$W^-(Y_1, Y_2|U_1) = \frac{1}{2}\sum_{u_2} W_1(y_1|u_1 \oplus u_2)W_2(y_2|u_2)$$

$$W^+(Y_1, Y_2, U_1|U_2) = \frac{1}{2}W_1(y_1|u_1 + u_2)W_2(y_2|u_2)$$

Polarization of general channels

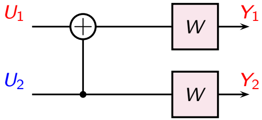


$$W^-(Y_1, Y_2|U_1) = \frac{1}{2}\sum_{u_2} W_1(y_1|u_1 \oplus u_2)W_2(y_2|u_2)$$

$$W^+(Y_1, Y_2, U_1|U_2) = \frac{1}{2}W_1(y_1|u_1 + u_2)W_2(y_2|u_2)$$

$$I(W^-) + I(W^+) = I(W) + I(W) = 2I(W)$$

Polarization of general channels



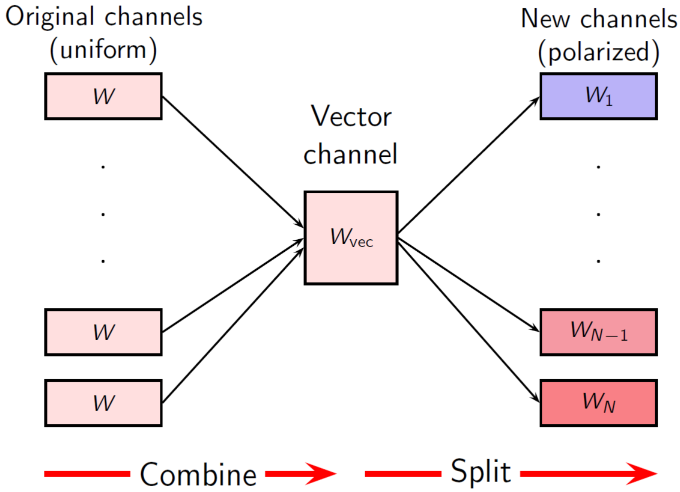
$$W^-(Y_1, Y_2|U_1) = \frac{1}{2}\sum_{u_2} W_1(y_1|u_1 \oplus u_2)W_2(y_2|u_2)$$

$$W^+(Y_1, Y_2, U_1|U_2) = \frac{1}{2}W_1(y_1|u_1 + u_2)W_2(y_2|u_2)$$

$$I(W^-) + I(W^+) = I(W) + I(W) = 2I(W)$$

- **Mrs Gerber's Lemma:** If $I(W) = 1 - \mathcal{H}(p)$, then $I(W^+) - I(W^{-1}) \geq 2\mathcal{H}(2p(1-p)) - \mathcal{H}(p)$

General Polar Construction



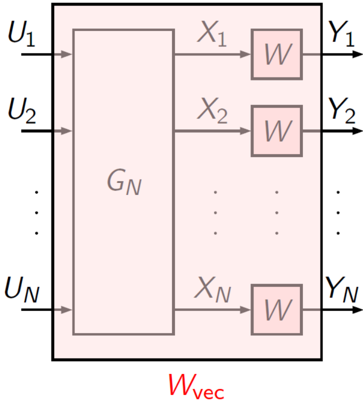
General Polar Construction

- Begin with N copies of W,
- use a 1-1 mapping

$$G_N : \{0,1\}^N \to \{0,1\}^N$$

- to create a vector channel

$$W_{\text{vec}} : U^N \to Y^N$$



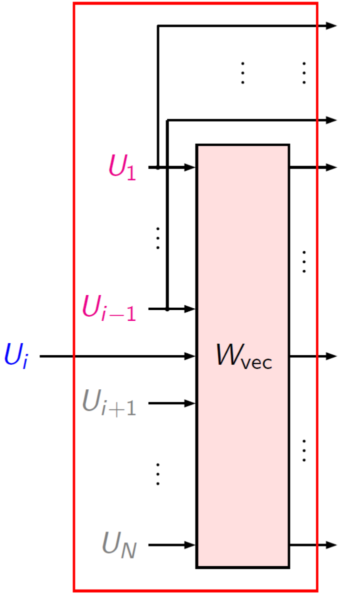
General Polar Construction

Splitting

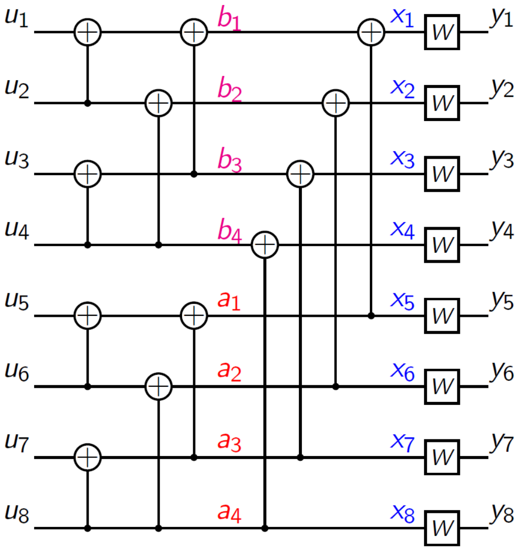
$$C(W_{\text{vec}}) = I(U^N; Y^N)$$
$$= \sum_{i=1}^{N} I(U_i; Y^N, U^{i-1})$$
$$= \sum_{i=1}^{N} C(W_i)$$

Define bit-channels

$$W_i : U_i \to (Y^N, U^{i-1})$$

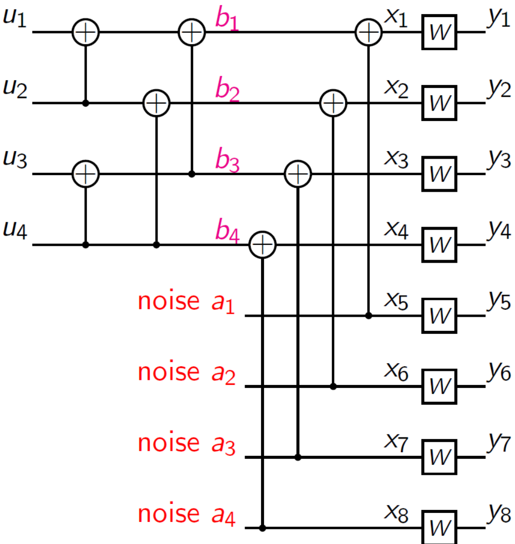


Successive Cancellation Decoder



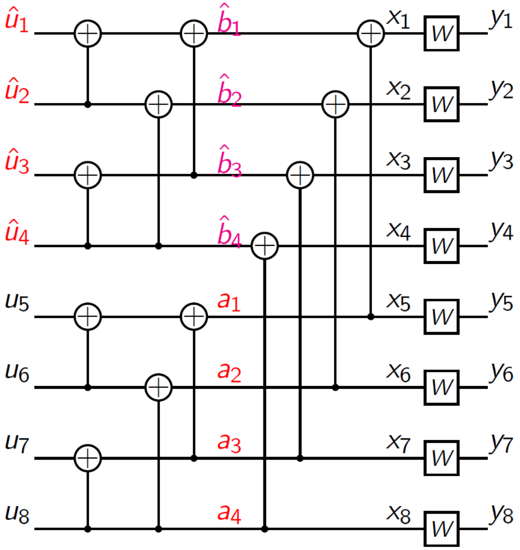
Successive Cancellation Decoder

First phase: treat **a** as noise, decode (u_1, u_2, u_3, u_4)



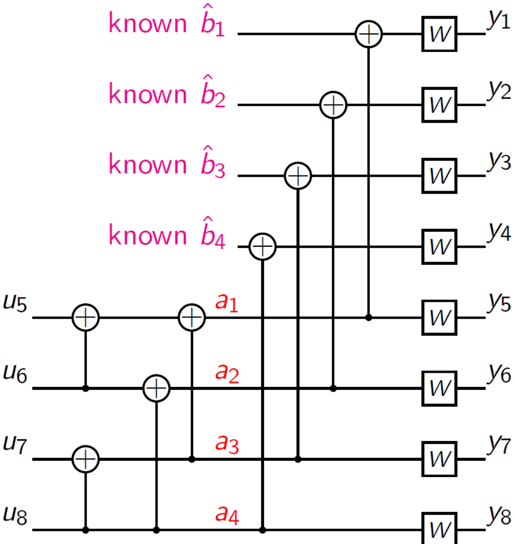
Successive Cancellation Decoder

End of first phase



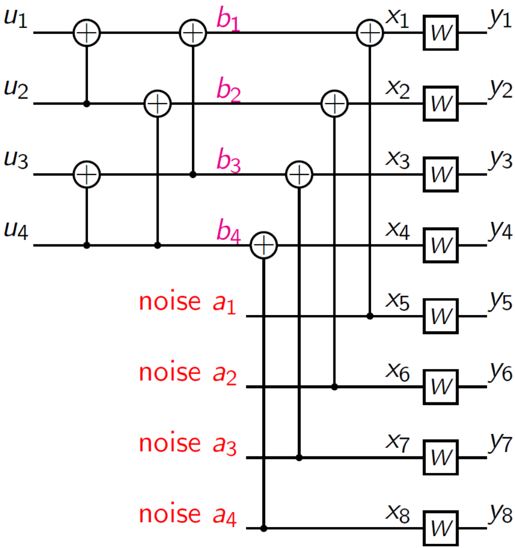
Successive Cancellation Decoder

Second phase: Treat $\hat{\mathbf{b}}$ as known, decode $(u_5, u_6, u_7, u_8$



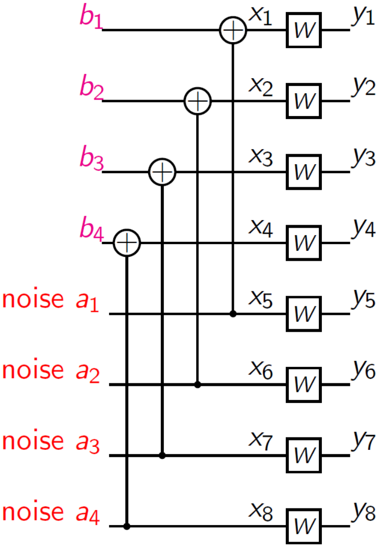
Successive Cancellation Decoder

First phase in detail



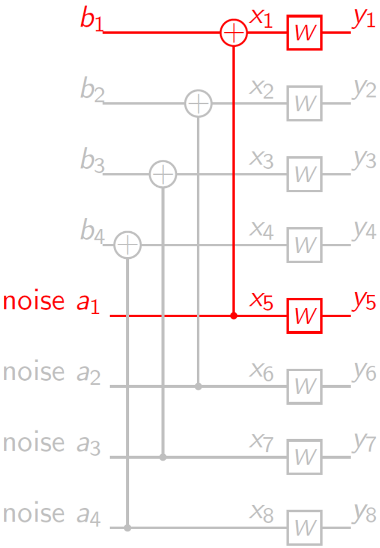
Successive Cancellation Decoder

Equivalent channel model



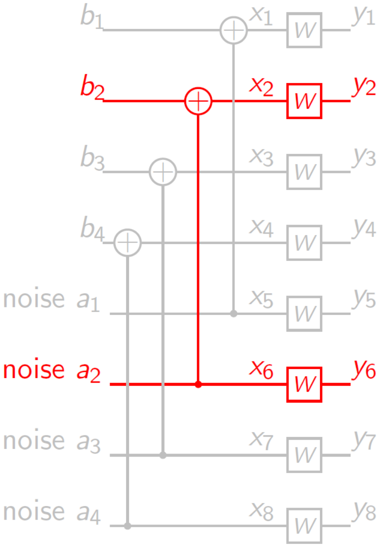
Successive Cancellation Decoder

First copy of W^-



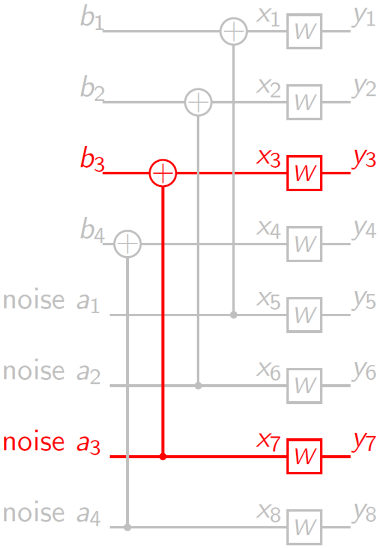
Successive Cancellation Decoder

Second copy of W^-



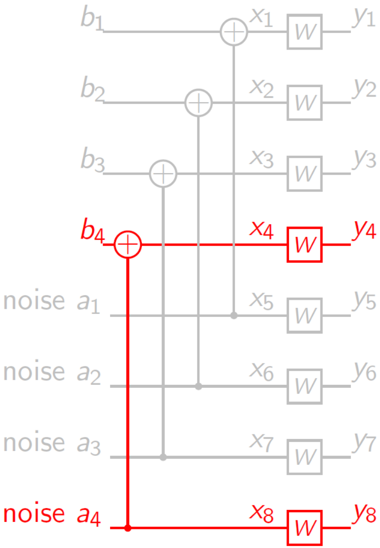
Successive Cancellation Decoder

Third copy of W^-



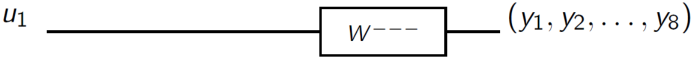
Successive Cancellation Decoder

Fourth copy of W^-



Successive Cancellation Decoder

Decoding on W^{---}



Compute

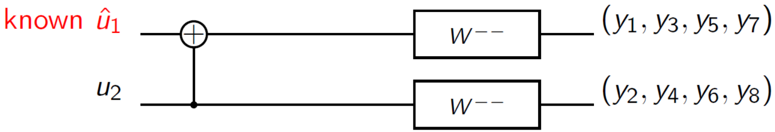
$$L^{---} \triangleq \frac{W^{---}(y_1, \ldots, y_8 \mid u_1 = 0)}{W^{---}(y_1, \ldots, y_8 \mid u_1 = 1)}.$$

Set

$$\hat{u}_1 = \begin{cases} u_1 & \text{if } u_1 \text{ is frozen} \\ 0 & \text{else if } L^{---} > 0 \\ 1 & \text{else} \end{cases}$$

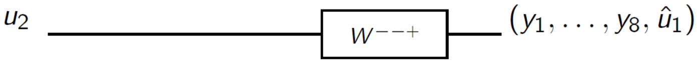
Successive Cancellation Decoder

Decoding on W^{--+}



Successive Cancellation Decoder

Decoding on W^{--+}



Compute

$$L^{--+} \triangleq \frac{W^{--+}(y_1, \ldots, y_8, \hat{u}_1 \mid u_2 = 0)}{W^{--+}(y_1, \ldots, y_8, \hat{u}_1 \mid u_2 = 1)}.$$

Set

$$\hat{u}_2 = \begin{cases} u_2 & \text{if } u_2 \text{ is frozen} \\ 0 & \text{else if } L^{--+} > 0 \\ 1 & \text{else} \end{cases}$$

Polar Coding Theorem

Theorem

For any rate $R < I(W)$ and block-length N, the probability of frame error for polar codes under successive cancelation decoding is bounded as

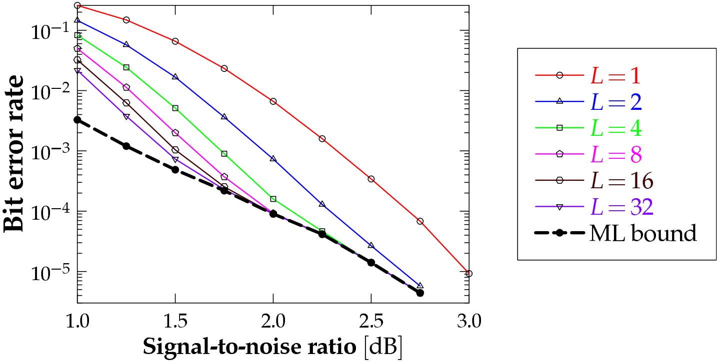
$$P_e(N, R) = o\left(2^{-\sqrt{N}+o(\sqrt{N})}\right)$$

Improved decoders

- List decoder (Tal and Vardy, 2011)

  First produce L candidate decisions
  Pick the most likely word from the list
  Complexity $O(LN \log N)$

List decoder



Polar Coding Summary

Summary

Given W, $N = 2^n$, and $R < I(W)$, a polar code can be constructed such that it has

- construction complexity $\mathcal{O}(N\text{poly}(log(N)))$,
- encoding complexity $\approx N \log N$,
- successive-cancellation decoding complexity $\approx N \log N$,
- frame error probability $P_e(N, R) = o\left(2^{-\sqrt{N}+o(\sqrt{N})}\right)$.

5G Communications

- The jump from 4G to 5G is far larger than any previous jumps–from 2G to 3G; 3G to 4G
- The global 5G market is expected reach a value of **251 Bn** by 2025

5G Communications

- The jump from 4G to 5G is far larger than any previous jumps–from 2G to 3G; 3G to 4G
- The global 5G market is expected reach a value of **251 Bn** by 2025
- In 2016, **27 Gbps** downlink speed was reached using Polar Codes!
- Current LTE download speed is **5-12 Mbps**

5G Communications

- The jump from 4G to 5G is far larger than any previous jumps–from 2G to 3G; 3G to 4G
- The global 5G market is expected reach a value of **251 Bn** by 2025
- In 2016, **27 Gbps** downlink speed was reached using Polar Codes!
- Current LTE download speed is **5-12 Mbps**
- In November 2016, 3GPP agreed to adopt Polar codes for control channels in 5G. LDPC codes will also be used in data channels.

References

- E. Arikan, Channel Polarization: A Method for Constructing Capacity-Achieving Codes for Symmetric Binary-Input Memoryless Channels, IEEE IT, 2009
- E. Arikan, Polar Coding Tutorial, Simons Institute, UC Berkeley, 2015
- B.C. Geiger, The Fractality of Polar and Reed–Muller Codes, Entropy, 2018