ISIT 2017 | Emre Telatar | The Flesh of Polar Codes | 2017-06-29

The flesh of polar codes

Emre Telatar

EPFL

Aachen — June 29, 2017

0:26 / 57:55

Easy channels

Among all channels, there are two classes for which it is easy to communicate optimally:

- The perfect channels: the output Y determines the input X.
- The useless channels: the output Y is independent of the input X.

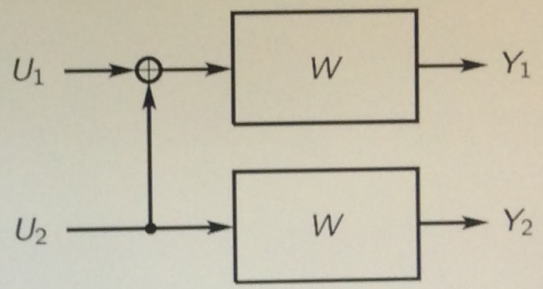
In a perfect world all channels would be of one of those extremal types.

Arıkan's polarization is a technique to convert any binary-input channel to a mixture of binary-input extremal channels.

- The technique is information lossless, and of low complexity.

Polar transform

Given two copies of a binary input channel $W: \mathbb{F}_2 \to \mathcal{Y}$



- Set

$$X_1 = U_1 \oplus U_2$$
$$X_2 = U_2$$

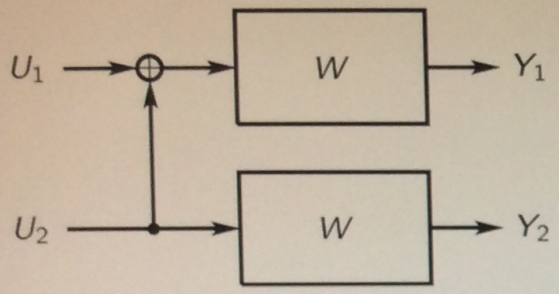
with (U_1, U_2) uniform on $\mathbb{F}_2^2$.

- As (X_1, X_2) is also uniform on $\mathbb{F}_2^2$,

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

- $I(W^-) \le I(W) \le I(W^+)$.

Are the synthetic channels real?



The students need to be convinced they are not being fooled.

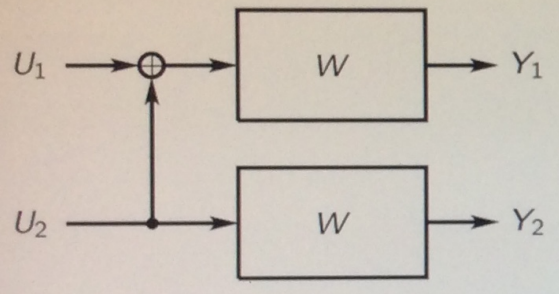
- The channel $W^- : U_1 \to Y_1 Y_2$ is a bonafide channel: U_1 is controlled by the transmitter, $Y_1 Y_2$ is observed by the receiver.
- The channel $W^+ : U_2 \to Y_1 Y_2 U_1$ poses problems: U_1 is not observed by the receiver.

- We can synthesize W^+ by first processing the output of W^- to estimate U_1. Even then, we only have $\hat{U}_1$. Is it legitimate to pretend that we have U_1 instead of $\hat{U}_1$?

Genie-aided receiver:		Implementable receiver:
$\tilde{U}_1 = \phi_1(Y^2)$	vs	$\hat{U}_1 = \phi_1(Y^2)$
$\tilde{U}_2 = \phi_2(Y^2 U_1)$		$\hat{U}_2 = \phi_2(Y^2 \hat{U}_1)$

- If the genie-aided receiver makes no errors, then, the implementable receiver also makes no errors. Indeed, the block error events $\{\tilde{U}^2 \neq U^2\}$ and $\{\hat{U}^2 \neq U^2\}$ are identical.
- We can choose to analyze the genie-aided receiver instead of the implementable one. For the purposes of this analysis the synthetic channels are real. The implementation will use $\hat{U}_1$ instead.

Example: Erasure channel



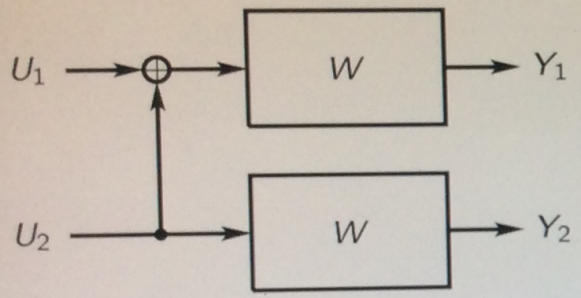
Suppose W is a BEC(p), i.e.,

$$Y = \begin{cases} X & \text{with probability } 1-p, \\ ? & \text{with probability } p \end{cases}$$

- W^- has input U_1, output

$$(Y_1, Y_2) = \begin{cases} (U_1 \oplus U_2, U_2) & \text{w.p. } (1-p)^2 \\ (\ ? \ , U_2) & \text{w.p. } p(1-p) \\ (U_1 \oplus U_2, \ ?\ ) & \text{w.p. } (1-p)p \\ (\ ? \ , \ ?\ ) & \text{w.p. } p^2 \end{cases}$$

Example: Erasure channel



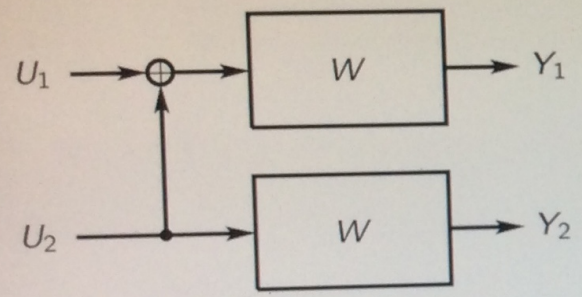
Suppose W is a BEC(p), i.e.,

$$Y = \begin{cases} X & \text{with probability } 1-p, \\ ? & \text{with probability } p \end{cases}$$

- W^- is a BEC(p^-); $p^- = 2p - p^2$.
- W^+ has input U_2, output

$$(Y_1, Y_2, U_1) = \begin{cases} (U_1 \oplus U_2, U_2, U_1) & \text{w.p. } (1-p)^2 \\ (\ ? \ , U_2, U_1) & \text{w.p. } p(1-p) \\ (U_1 \oplus U_2, \ ? \ , U_1) & \text{w.p. } (1-p)p \\ (\ ? \ , \ ? \ , U_1) & \text{w.p. } p^2 \end{cases}$$

Example: Erasure channel



Suppose W is a BEC(p), i.e.,

$$Y = \begin{cases} X & \text{with probability } 1-p, \\ ? & \text{with probability } p \end{cases}$$

- W^- is a BEC(p^-); $p^- = 2p - p^2$.
- W^+ is a BEC(p^+); $p^+ = p^2$.

- We already begin to see some extremalization: W^+ is better than W, while W^- is worse.

Polarization construction

What we can do once, we can do many times. Given $W : \mathbb{F}_2 \to \mathcal{Y}$,

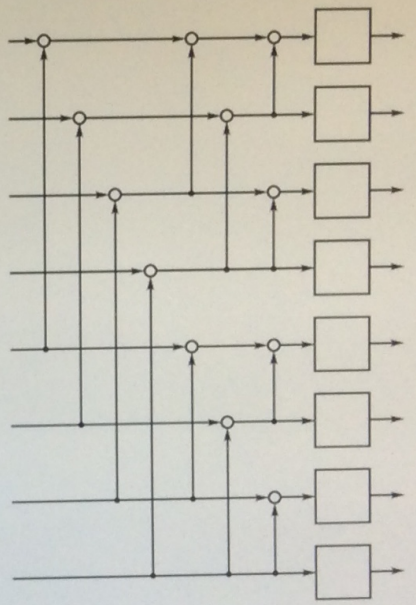
- Duplicate $W : X \to Y$ and obtain $W^- : U_1 \to Y_1 Y_2$ and $W^+ : U_2 \to Y_1 Y_2 U_1$.
- Duplicate W^- (and also W^+)
- and obtain

$$W^{--} : V_1 \to Y_1 Y_2 Y_3 Y_4$$
$$W^{-+} : V_2 \to Y_1 Y_2 Y_3 Y_4 V_1$$

and also

$$W^{+-} : V_3 \to Y_1 Y_2 Y_3 Y_4 V_1 V_2$$
$$W^{++} : V_4 \to Y_1 Y_2 Y_3 Y_4 V_1 V_2 V_3$$

- Duplicate W^{--} (and W^{-+}, W^{+-}, W^{++}) and obtain W^{---} and W^{--+} (and W^{-+-}, W^{-++}, W^{+--}, W^{+-+}, W^{++-}, W^{+++}).

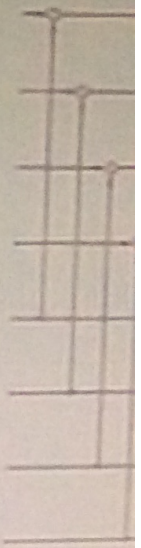


ISIT 2017 | Emre Telatar | The Flesh of Polar Codes | 2017-06-29

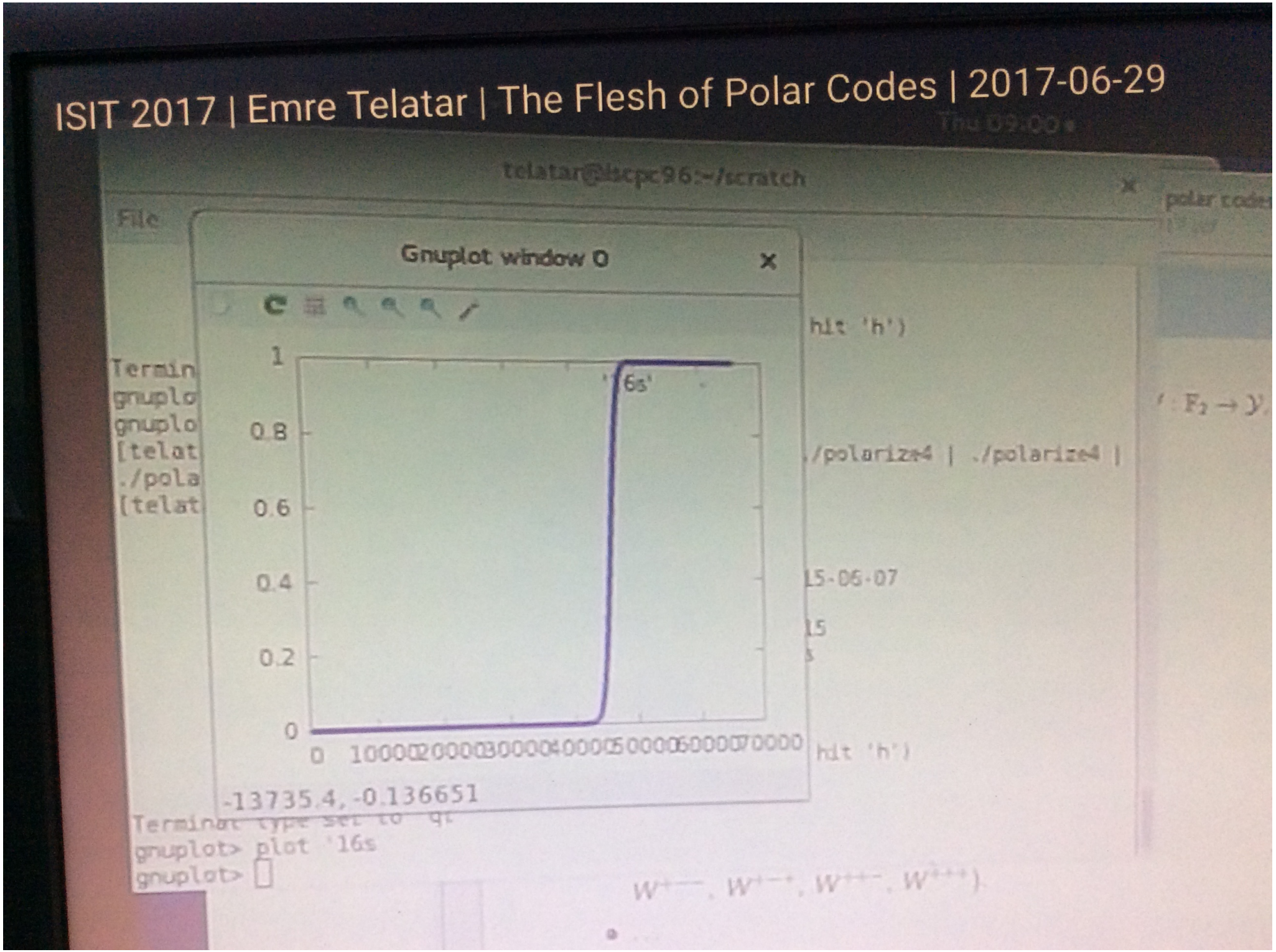
telatar@iscpc96:~/scratch

File Edit View Search Terminal Help

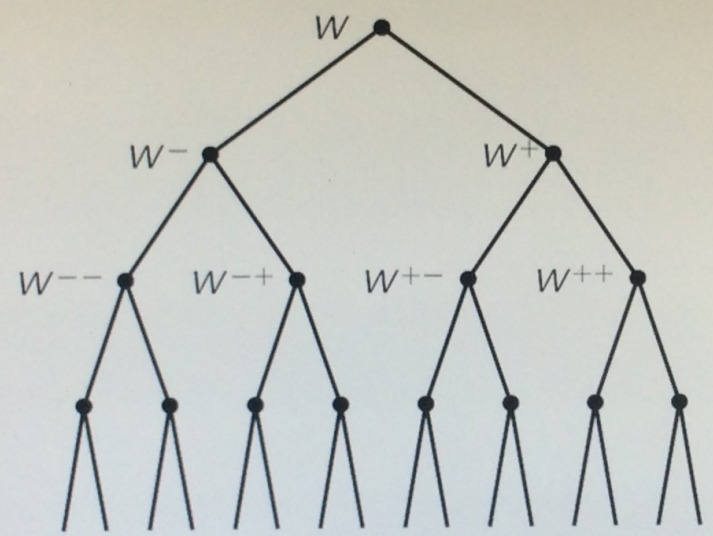
```
0.09
0.51
.5
0.25
0.75
[telatar@iscpc96 scratch]$ echo 0.3 | ./polarize | ./polarize | ./polarize | ./p
olarize
4.30467e-09
0.000131216
0.000260319
0.0320085
0.000873179
0.058226
0.0987531
0.529747
0.00457679
0.130727
0.2048
0.700296
0.333446
0.82145
0.888027
0.996677
[telatar@iscpc96 scratch]$
```



ISIT 2017 | Emre Telatar | The Flesh of Polar Codes | 2017-06-29



Family tree of channels



Polarization happens

Theorem

For any $\epsilon > 0$ and any W, the fraction of ϵ-mediocre channels vanishes as we repeatedly apply the polar transform:

$$\mu_t(\epsilon) := \frac{1}{2^t} \sum_{s^t \in \{+,-\}^t} \mathbb{1}\{I(W^{s^t}) \in (\epsilon, 1-\epsilon)\}, \quad \lim_{t\to\infty} \mu_t(\epsilon) = 0.$$

Theorem (Restricted to BEC)

For any $\epsilon > 0$, and any $W = \mathrm{BEC}(p)$, the fraction of ϵ-mediocre channels vanishes as we repeatedly apply the polar transform:

$$\mu_t(\epsilon) := \frac{1}{2^t} \sum_{s^t \in \{+,-\}^t} \mathbb{1}\{p(s^t) \in (\epsilon, 1-\epsilon)\}, \quad \lim_{t\to\infty} \mu_t(\epsilon) = 0.$$

Proof of polarization for BECs

Theorem (Restricted to BEC)

For any $\epsilon > 0$, and any $W = \mathrm{BEC}(p)$, the fraction of ϵ-mediocre channels vanishes as we repeatedly apply the polar transform:

$$\mu_t(\epsilon) := \frac{1}{2^t} \sum_{s^t \in \{+,-\}^t} \mathbb{1}\{p(s^t) \in (\epsilon, 1-\epsilon)\}, \quad \lim_{t\to\infty} \mu_t(\epsilon) = 0.$$

- For a channel $P = \mathrm{BEC}(q)$ define its ugliness, $\mathrm{ugly}(P) = \sqrt{4q(1-q)}$.
- The children of P, the channels P^+ and P^-, have ugliness

$$\begin{aligned} \mathrm{ugly}(P^+) &= \sqrt{4q^2(1-q^2)} \\ &= \mathrm{ugly}(P)\sqrt{q(1+q)} \end{aligned} \qquad \begin{aligned} \mathrm{ugly}(P^-) &= \sqrt{4q(2-q)(1-q)^2} \\ &= \mathrm{ugly}(P)\sqrt{(2-q)(1-q)} \end{aligned}$$

- $\frac{1}{2}\mathrm{ugly}(P^+) + \frac{1}{2}\mathrm{ugly}(P^-) = \mathrm{ugly}(P)\frac{1}{2}\left[\sqrt{q(1+q)} + \sqrt{(2-q)(1-q)}\right] \le \mathrm{ugly}(P)\sqrt{\frac{3}{4}}$.
- (observation first made by B. Hajek)

Polarization proof for BECs

Theorem (Restricted to BEC)

For any $\epsilon > 0$, and any $W = \mathrm{BEC}(p)$, the fraction of ϵ-mediocre channels vanishes as we repeatedly apply the polar transform:

$$\mu_t(\epsilon) := \frac{1}{2^t} \sum_{s^t \in \{+,-\}^t} \mathbb{1}\{p(s^t) \in (\epsilon, 1-\epsilon)\}, \quad \lim_{t\to\infty} \mu_t(\epsilon) = 0.$$

- Recall: $\mathrm{ugly}(\mathrm{BEC}(q)) := \sqrt{4q(1-q)}$.
- Recall: $\frac{1}{2}[\mathrm{ugly}(P^+) + \mathrm{ugly}(P^-)] \le \sqrt{\frac{3}{4}}\,\mathrm{ugly}(P)$.
- Thus: $\frac{1}{2^t} \sum_{s\in\{+,-\}^t} \mathrm{ugly}(W^s) \le \left(\frac{3}{4}\right)^{t/2} \mathrm{ugly}(W)$.
- Note: $\mathbb{1}\{p(s) \in (\epsilon, 1-\epsilon)\} \le \mathrm{ugly}(W^s) \,/\, \sqrt{4\epsilon(1-\epsilon)}$
- Hence: $\mu_t(\epsilon) \le \left(\frac{3}{4}\right)^{t/2} \,/\, \sqrt{4\epsilon(1-\epsilon)}$
- QED.

...

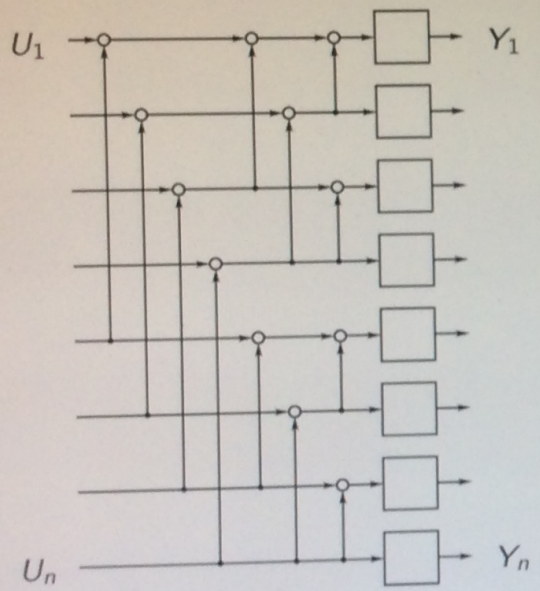
Corollary

For any $\epsilon \in (0,1)$, for any $W = \mathrm{BEC}(p)$, the fraction of ϵ-good channels approaches $1-p$:

$$\gamma_t(\epsilon) := \frac{1}{2^t} \sum_{s \in \{+,-\}^t} \mathbb{1}\{p(s) \leq \epsilon\}; \quad \lim_{t \to \infty} \gamma_t(\epsilon) = 1 - p.$$

Polar coding

This suggests the following coding scheme: given $W = \mathrm{BEC}(p)$, $R < 1 - p$,



- Start with $n = 2^t$ copies of W. (I.e., blocklength is 2^t.)
- Apply the polar transform for t generations to synthesize the n channels $W^{+\cdots+}, \ldots, W^{-\cdots-}$.
- Set the inputs of the best nR channels to uncoded data. Freeze the inputs of the remaining channels to 0.
- Number of ϵ-good channels is $\approx (1-p)n$, so all the nR channels are good.
- At the receiver, successively decode $U_1, \ldots, U_n$.
- The block error probability is upper bounded by

$$\sum_{s:p(s)\leq\epsilon} p(s) \leq n\epsilon$$

- Hmmm.

Hmmm.

The polarization theorem we just proved is not strong enough. We need to be able to choose $\epsilon = \epsilon_t \ll 1/n$, and still have polarization ($n = 2^t$).

Theorem

For any $\delta > 0$, and any $W = \mathrm{BEC}(p)$, with $n = 2^t$, and $\epsilon_t := \exp_2(-n^{\frac{1}{2}-\delta})$, the fraction of ϵ_t-good channels approaches $1 - p$:

$$\lim_{t \to \infty} \gamma_t(\epsilon_t) = 1 - p.$$

Taintedness

To prove this stronger polarization statement we need two concepts:

- Call a channel $W^s = W^{s_1 \cdots s_t}$ at generation t to be ϵ-tainted if it has an ϵ-mediocre ancestor $W^{s_1 \cdots s_i}$ at any generation $i \in [\sqrt{t}, t]$.
- The fraction of ϵ-tainted channels at generation t is upper bounded by

$$\sum_{i=\sqrt{t}}^{t} \mu_i(\epsilon) \leq \frac{1}{\sqrt{4\epsilon(1-\epsilon)}} \sum_{i=\sqrt{t}}^{t} \left(\tfrac{3}{4}\right)^{i/2} \leq \mathrm{const}(\epsilon) \left(\tfrac{3}{4}\right)^{\sqrt{t}/2}$$

  which vanishes as t gets large.
- For small ϵ, the transformations $q \to q^+$ and $q \to q^-$ don't jump between $[0, \epsilon]$ and $[1-\epsilon, 1]$.
- So if ϵ is small, and if $W^{s_1 \cdots s_t}$ is ϵ-untainted, then, all its ancestors $\{W^{s^i} : i \in [\sqrt{t}, t]\}$ are the same type: either all are good (BEC(q) with $q \leq \epsilon$) or all are bad (with $q \geq 1-\epsilon$).

Unluckyness

The other notion we need is that of luck:

- Call a channel $W^s = W^{s_1 \cdots s_t}$ at generation t to be ϵ-unlucky if $\{s_i : i \in [\sqrt{t}, t]\}$ does not have its fair share of $+$'s:

$$\sum_{i=\sqrt{t}}^{t} \mathbb{1}\{s_i = +\} < \left(\tfrac{1}{2} - \epsilon\right) t.$$

- The fraction of ϵ-unlucky channels at generation t vanishes as t gets large.

...

Consider the t'th generation descendants of $W = \mathrm{BEC}(p)$, pick a small $\epsilon > 0$.

- The fraction of ϵ-good, ϵ-untainted, ϵ-lucky channels approaches $1 - p$ as t gets large. Take any such channel, and trace its ancestors in generations $[\sqrt{t}, t]$. For each generational step

$$\text{either} \quad p(s^{i+1}) = p(s^i)^2$$
$$\text{or} \quad p(s^{i+1}) = p(s^i)(2 - p(s^i)) \le p(s^i)\,2 = p(s^i)\,\epsilon^{-\epsilon'} \le p(s^i)^{1-\epsilon'} \qquad \text{with } \epsilon' = 1/\log_2(1/\epsilon)$$

- Taking logs twice

$$\log_2 \log_2 \frac{1}{p(s^{i+1})} \ge \log_2 \log_2 \frac{1}{p(s^i)} + \begin{cases} 1 & s_{i+1} = + \\ -\epsilon'' & s_{i+1} = - \end{cases} \qquad \text{with } \epsilon'' = -\log_2(1 - \epsilon')$$

- Chaining the inequalities from $i = \sqrt{t}$ till $t - 1$,

$$\log_2 \log_2 \frac{1}{p(s^t)} \ge \log_2 \log_2 \frac{1}{\epsilon} + \left(\tfrac{1}{2} - \epsilon\right)t - t\epsilon'' \ge \left(\tfrac{1}{2} - \delta\right)t \quad \text{for small enough } \epsilon$$
$$-\log_2 p(s^t) \ge 2^{(1/2-\delta)t} = n^{1/2-\delta}.$$

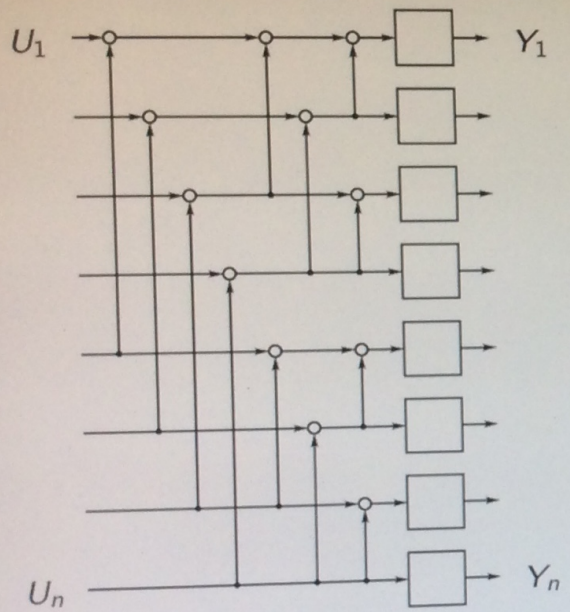
...

Corollary

Polar codes are capacity achieving with error probability exponentially small in the square root of the block length.

Given $\delta > 0$, $W = \mathrm{BEC}(p)$, and $R < 1 - p$; for large $n = 2^t$ the polar code constructed by sending data on the nR best synthetic channels has error probability at most $n2^{-n^{1/2-\delta}}$.

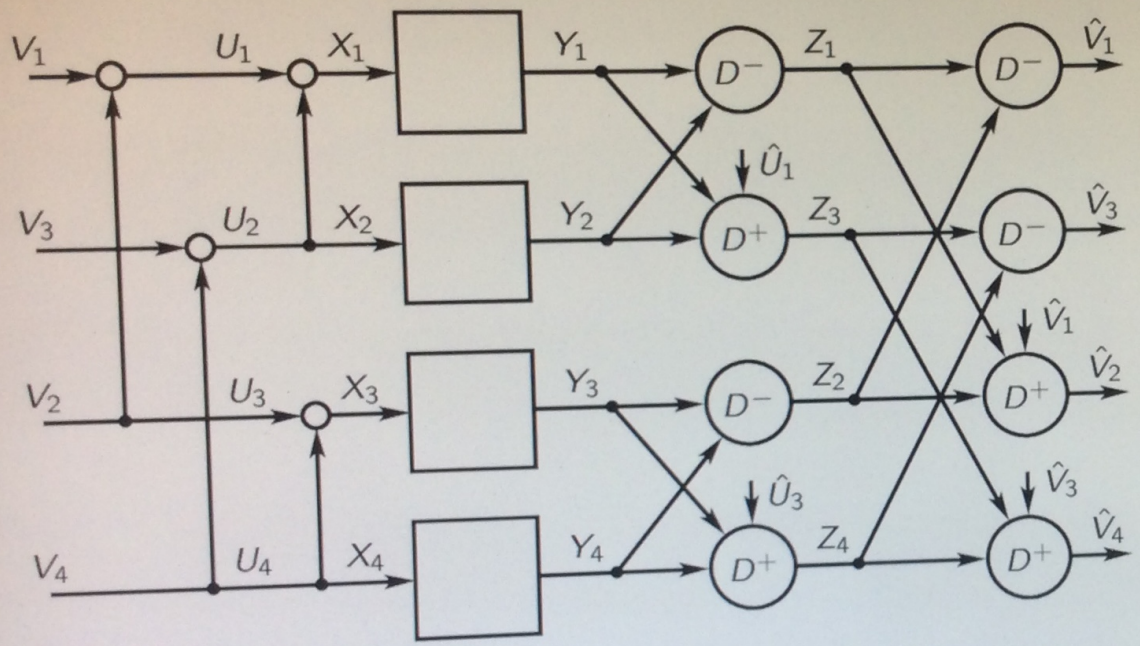
Encoding Complexity



For a polar code of blocklength $n = 2^t$,

- There are t encoding stages
- Each stage with $n/2$ exclusive-or operations.
- Encoding complexity is $n \log_2 n$.

Decoding Complexity



Receiver's first task: construct the output of the BEC W^-. Recall $(Y_1, Y_2) = (U_1 \oplus U_2, U_2)$ or useless.

$$D^-(y_1, y_2) = \begin{cases} y_1 \oplus y_2 & y_1 \neq ?, y_2 \neq ? \\ ? & \text{else} \end{cases}$$

$$D^+(z_1, z_2, v_1) = \begin{cases} z_2 & z_2 \neq ? \\ z_1 \oplus v_1 & z_1 \neq ? \\ ? & \text{else} \end{cases}$$

We have $n \log_2 n$ computation units each executing once. Decoding complexity is $n \log_2 n$.

Remarks

- Polar codes are easy to teach, and the students appear to appreciate hearing about them.
- They provide a constructive argument to prove the channel coding theorem, and allow a symmetry-in-teaching with source coding.
- See [Alsan & T., A simple proof of polarization and polarization for non-stationary memoryless channels, T-IT 2016] for a second moment method to prove the first polarization theorem for general binary input channels.
- I. Tal has an arXiv paper that also describes a simple proof of the second polarization theorem for general binary input channels.
- Do read Arıkan's paper on polarization, and also his 'On the origin of polar coding' for further insight on how they came to be.