# Polarization and polar codes

Emre Telatar

EPFL

# Extremal Channels

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

# Extremal 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 $X$.

# Extremal 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 $X$.

# Extremal 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 $X$.

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

# Extremal 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 $X$.

Arıkan's polar coding 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⋆.

# Extremal Channels

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

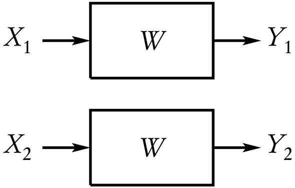
- The perfect channels: the output $Y$ determines the input $X$.
- The useless channels: the output $Y$ is independent of the input $X$.

Arıkan's polar coding 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$^{\star}$.
- I am not the inventor of this technique.

# Building block

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



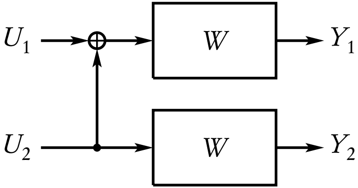
# Building block

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

- Set

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

with $U_1, U_2$ i.i.d., uniform on $\mathbb{F}_2$.



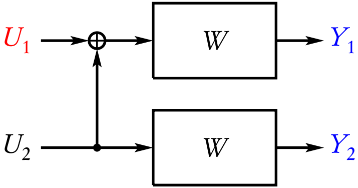
# Building block

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

- Set

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

with $U_1, U_2$ i.i.d., uniform on $\mathbb{F}_2$.

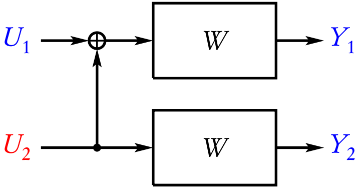



- This induces two synthetic channels $W^- : \mathbb{F}_2 \to \mathscr{Y}^2$

# Building block

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

- Set
$$X_1 = U_1 \oplus U_2$$
$$X_2 = U_2$$
with $U_1, U_2$ i.i.d., uniform on $\mathbb{F}_2$.




- This induces two synthetic channels $W^- : \mathbb{F}_2 \to \mathscr{Y}^2$ and $W^+ : \mathbb{F}_2 \to \mathscr{Y}^2 \times \mathbb{F}_2$.

# Building block

Note that

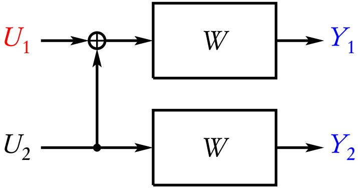
$$W^{-}(y_1, y_2 | u_1) = \sum_{u_2 \in \mathbb{F}_2} \tfrac{1}{2} W(y_1 | u_1 \oplus u_2) W(y_2 | u_2)$$

$$W^{+}(y_1, y_2, u_1 | u_2) = \tfrac{1}{2} W(y_1 | u_1 \oplus u_2) W(y_2 | u_2)$$

# Building block

Properties of $W \mapsto (W^-, W^+)$:

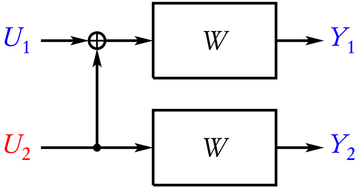
$$I(W^-) = I(U_1; Y_1 Y_2)$$



# Building block

Properties of $W \mapsto (W^-, W^+)$:

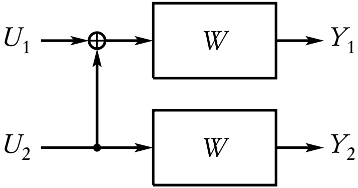
$$I(W^-) = I(U_1; Y_1 Y_2)$$
$$I(W^+) = I(U_2; Y_1 Y_2 U_1)$$



# Building block

Properties of $W \mapsto (W^-, W^+)$:

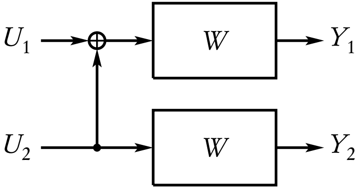
$$\begin{aligned} I(W^-) &= I(U_1; Y_1 Y_2) \\ I(W^+) &= I(U_2; Y_1 Y_2 U_1) \\ I(W^-) + I(W^+) &= I(U_1 U_2; Y_1 Y_2) \end{aligned}$$



# Building block

Properties of $W \mapsto (W^-, W^+)$:

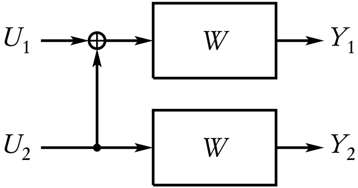
$$\begin{aligned} I(W^-) &= I(U_1; Y_1 Y_2) \\ I(W^+) &= I(U_2; Y_1 Y_2 U_1) \\ I(W^-) + I(W^+) &= I(U_1 U_2; Y_1 Y_2) \\ &= I(X_1 X_2; Y_1 Y_2) \end{aligned}$$



# Building block

Properties of $W \mapsto (W^-, W^+)$:

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

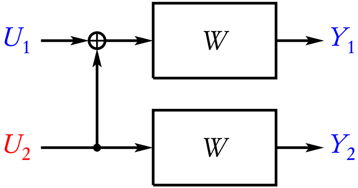



- $\frac{1}{2} I(W^-) + \frac{1}{2} I(W^+) = I(W)$.

# Building block

Properties of $W \mapsto (W^-, W^+)$:

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

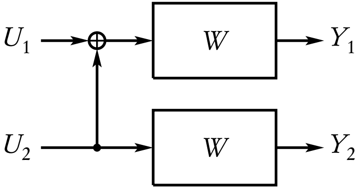



- $\frac{1}{2} I(W^-) + \frac{1}{2} I(W^+) = I(W)$.
- $I(W^+) \geq I(W)$

# Building block

Properties of $W \mapsto (W^-, W^+)$:

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




- $\frac{1}{2} I(W^-) + \frac{1}{2} I(W^+) = I(W)$.
- $I(W^+) \geq I(W) \geq I(W^-)$.

# Building block

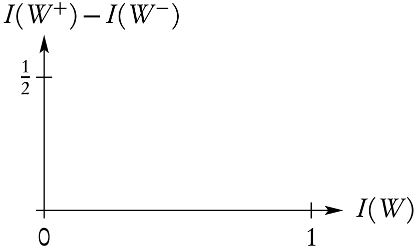
Properties of $W \mapsto (W^-, W^+)$:

- $\frac{1}{2}I(W^-) + \frac{1}{2}I(W^+) = I(W)$.
- $I(W^+) \geq I(W) \geq I(W^-)$.

# Building block

Properties of $W \mapsto (W^-, W^+)$:

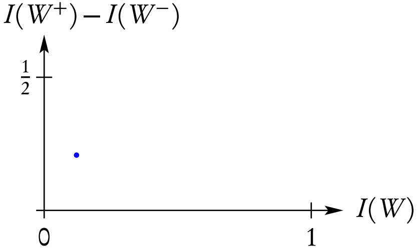
- $\frac{1}{2}I(W^-) + \frac{1}{2}I(W^+) = I(W)$.
- $I(W^+) \geq I(W) \geq I(W^-)$.



# Building block

Properties of $W \mapsto (W^-, W^+)$:

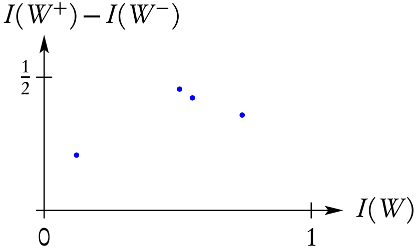
- $\frac{1}{2}I(W^-) + \frac{1}{2}I(W^+) = I(W)$.
- $I(W^+) \geq I(W) \geq I(W^-)$.



# Building block

Properties of $W \mapsto (W^-, W^+)$:

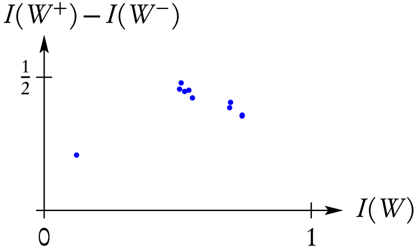
- $\frac{1}{2}I(W^-) + \frac{1}{2}I(W^+) = I(W)$.
- $I(W^+) \geq I(W) \geq I(W^-)$.



# Building block

Properties of $W \mapsto (W^-, W^+)$:

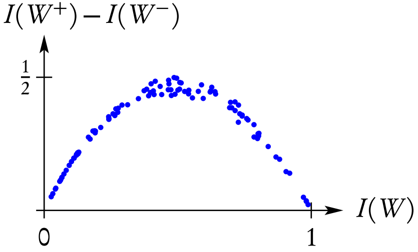
- $\frac{1}{2}I(W^-) + \frac{1}{2}I(W^+) = I(W)$.
- $I(W^+) \geq I(W) \geq I(W^-)$.



# Building block

Properties of $W \mapsto (W^-, W^+)$:

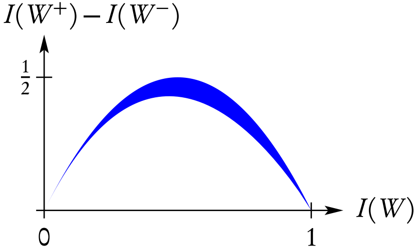
- $\frac{1}{2}I(W^-) + \frac{1}{2}I(W^+) = I(W)$.
- $I(W^+) \geq I(W) \geq I(W^-)$.



# Building block

Properties of $W \mapsto (W^-, W^+)$:

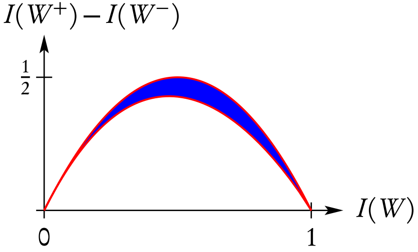
- $\frac{1}{2}I(W^-) + \frac{1}{2}I(W^+) = I(W)$.
- $I(W^+) \geq I(W) \geq I(W^-)$.



# Building block

Properties of $W \mapsto (W^-, W^+)$:

- $\frac{1}{2}I(W^-) + \frac{1}{2}I(W^+) = I(W)$.
- $I(W^+) \geq I(W) \geq I(W^-)$.



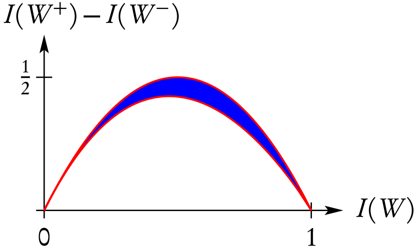
# Building block

Properties of $W \mapsto (W^-, W^+)$:

- $\frac{1}{2}I(W^-) + \frac{1}{2}I(W^+) = I(W)$.
- $I(W^+) \geq I(W) \geq I(W^-)$.
- For every $\epsilon > 0$ there is a $\delta > 0$ such that

  $$I(W^+) - I(W^-) < \delta \text{ implies}$$

  $$I(W) \notin (\epsilon, 1 - \epsilon).$$



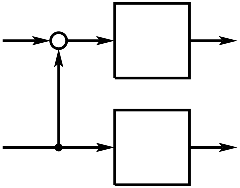
# Polarization construction

What we can do once, we can do many times.

# Polarization construction

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

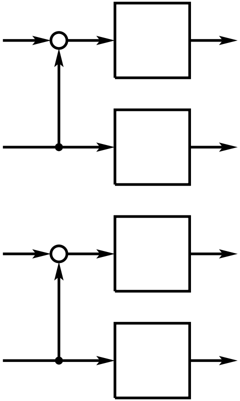
- Duplicate $W$ and obtain $W^-$ and $W^+$.



# Polarization construction

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

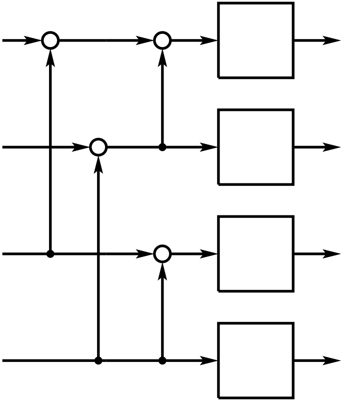
- Duplicate $W$ and obtain $W^-$ and $W^+$.
- Duplicate $W^-$ (and $W^+$),



# Polarization construction

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

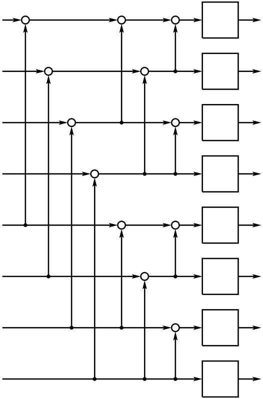
- Duplicate $W$ and obtain $W^-$ and $W^+$.
- Duplicate $W^-$ (and $W^+$),
- and obtain $W^{--}$ and $W^{-+}$ (and $W^{+-}$ and $W^{++}$).



# Polarization construction

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

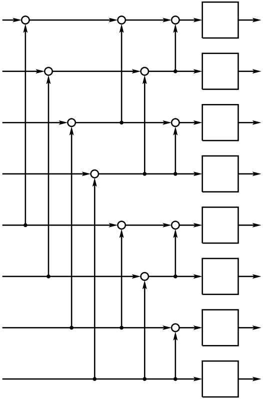
- Duplicate $W$ and obtain $W^-$ and $W^+$.
- Duplicate $W^-$ (and $W^+$),
- and obtain $W^{--}$ and $W^{-+}$ (and $W^{+-}$ and $W^{++}$).
- Duplicate $W^{--}$ (and $W^{-+}$, $W^{+-}$, $W^{++}$) and obtain $W^{---}$ and $W^{--+}$ (and $W^{-+-}$, $W^{-++}$, $W^{+--}$, $W^{+-+}$, $W^{++-}$, $W^{+++}$).



# Polarization construction

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

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



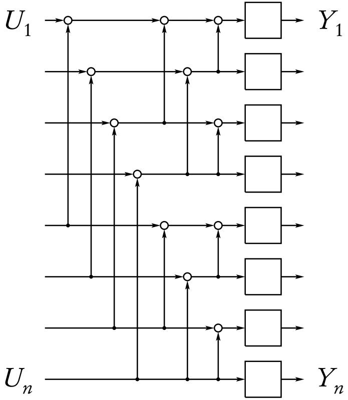
# Polarization construction

- $\ell$ levels into this process, we have transformed $n = 2^\ell$ uses of channel $W$ to one use each of $2^\ell$ channels

$$W^{b_1 \dots, b_\ell}, \quad b_j \in \{+, -\},$$

these are the channels

$$U_i \to Y^n U^{i-1}, \quad i = 1, \dots, n.$$



# Polarization construction

- $\ell$ levels into this process, we have transformed $n = 2^\ell$ uses of channel $W$ to one use each of $2^\ell$ channels

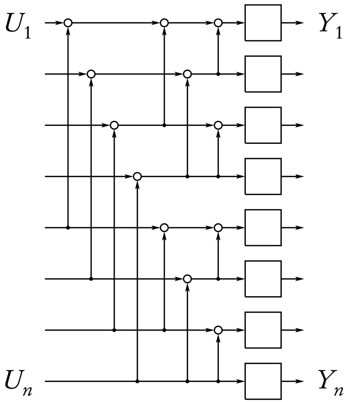
$$W^{b_1,\ldots,b_\ell}, \quad b_j \in \{+,-\},$$

these are the channels

$$U_i \to Y^n U^{i-1}, \quad i = 1,\ldots,n.$$

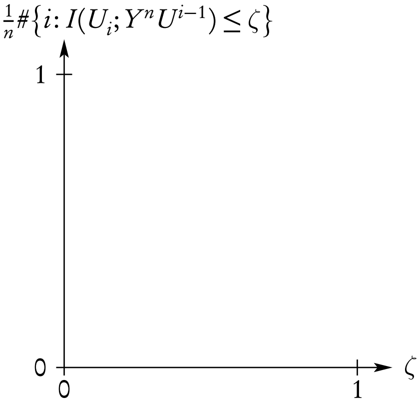
- The quantities $\{I(W^{b_1\ldots b_\ell})\}$ are exactly the $n$ quantities

$$I(U_i; Y^n U^{i-1}), \quad i = 1,\ldots,n$$



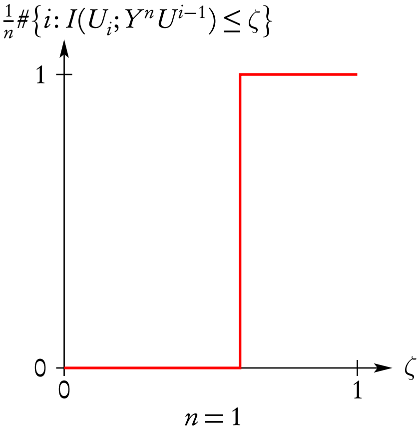
# Example

$W$ is a binary erasure channel, $p = 0.4$



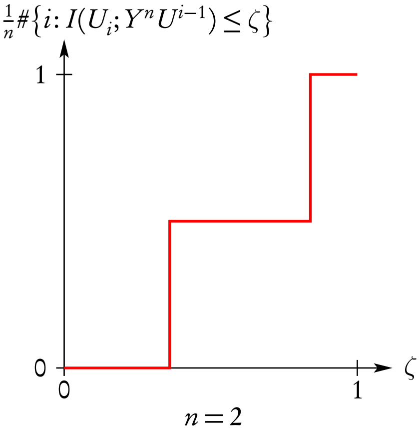
# Example

$W$ is a binary erasure channel, $p = 0.4$



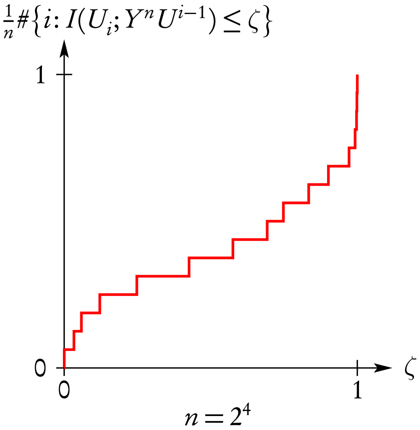
# Example

$W$ is a binary erasure channel, $p = 0.4$



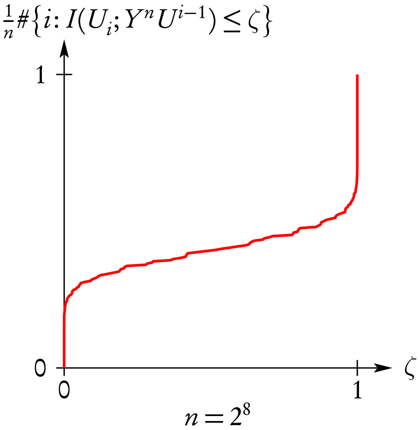
# Example

$W$ is a binary erasure channel, $p = 0.4$



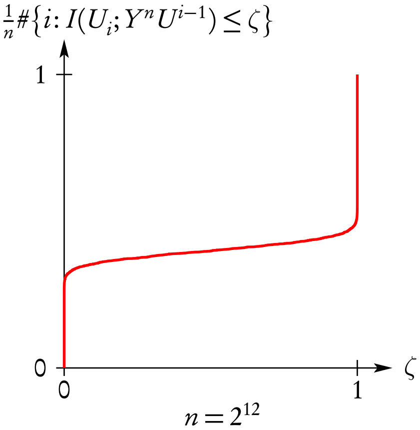
# Example

$W$ is a binary erasure channel, $p = 0.4$



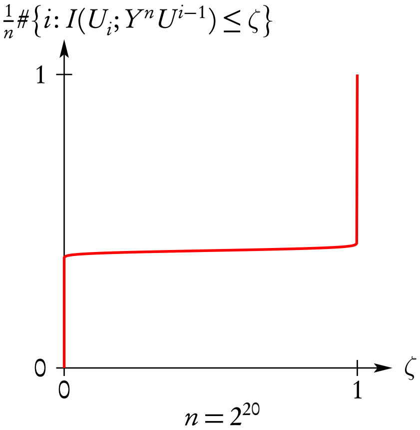
# Example

$W$ is a binary erasure channel, $p = 0.4$



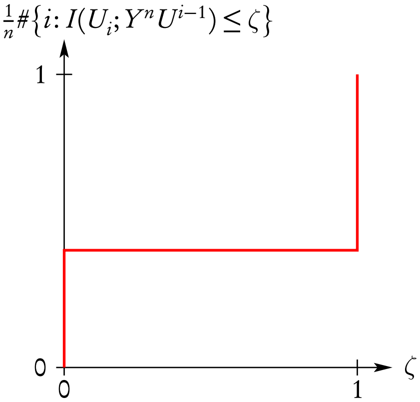
# Example

$W$ is a binary erasure channel, $p = 0.4$



# Example

$W$ is a binary erasure channel, $p = 0.4$



# Polarization

- Polarization refers to the phenomenon that in the limit, almost all channels are extremal, i.e.,

$$\frac{1}{n}\#\{i\colon I(U_i;Y^nU^{i-1})\in(\epsilon,1-\epsilon)\}\to 0$$

as $n=2^\ell$ gets large.

# Polarization

- Polarization refers to the phenomenon that in the limit, almost all channels are extremal, i.e.,

$$\frac{1}{n}\#\{i : I(U_i; Y^n U^{i-1}) \in (\epsilon, 1-\epsilon)\} \to 0$$

as $n = 2^\ell$ gets large.

- *If* this happens, the *good* limiting synthetic channels can be used to transmit uncoded data bits — the inputs to the other channels can be frozen to fixed values. Since the transformation is information lossless, the fraction of good channels (i.e., data rate) must be $I(W)$.

# Polarization

- Polarization refers to the phenomenon that in the limit, almost all channels are extremal, i.e.,

$$\frac{1}{n}\#\{i : I(U_i; Y^n U^{i-1}) \in (\epsilon, 1-\epsilon)\} \to 0$$

as $n = 2^\ell$ gets large.
- *If* this happens, the *good* limiting synthetic channels can be used to transmit uncoded data bits — the inputs to the other channels can be frozen to fixed values. Since the transformation is information lossless, the fraction of good channels (i.e., data rate) must be $I(W)$.
- The decoder can decode $U_1, U_2, \ldots, U_n$ successively.

# On successive decoding

The quantities $I(U_i; Y^n U^{i-1})$ are relevant for a genie-aided decoder:

$$\hat{U}_1 = \phi_1(Y^n)$$
$$\hat{U}_2 = \phi_2(Y^n, U_1)$$
$$\hat{U}_3 = \phi_3(Y^n, U^2)$$
$$\dots$$
$$\hat{U}_n = \phi_n(Y^n, U^{n-1})$$

# On successive decoding

The quantities $I(U_i; Y^n U^{i-1})$ are relevant for a genie-aided decoder:

$$\hat{U}_1 = \phi_1(Y^n)$$
$$\hat{U}_2 = \phi_2(Y^n, U_1)$$
$$\hat{U}_3 = \phi_3(Y^n, U^2)$$
$$\dots$$
$$\hat{U}_n = \phi_n(Y^n, U^{n-1})$$

vs

unaided decoder:

$$\hat{U}_1 = \phi_1(Y^n)$$
$$\hat{U}_2 = \phi_2(Y^n, \hat{U}_1)$$
$$\hat{U}_3 = \phi_3(Y^n, \hat{U}^2)$$
$$\dots$$
$$\hat{U}_n = \phi_n(Y^n, \hat{U}^{n-1}).$$

# On successive decoding

The quantities $I(U_i; Y^n U^{i-1})$ are relevant for a genie-aided decoder:

$$\hat{U}_1 = \phi_1(Y^n)$$
$$\hat{U}_2 = \phi_2(Y^n, U_1)$$
$$\hat{U}_3 = \phi_3(Y^n, U^2)$$
$$\dots$$
$$\hat{U}_n = \phi_n(Y^n, U^{n-1})$$

vs

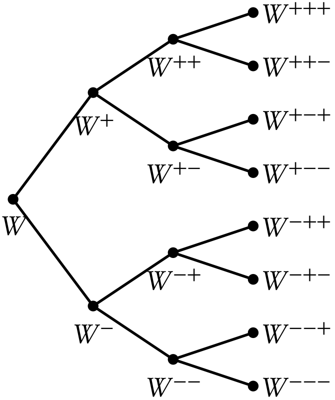
unaided decoder:

$$\hat{U}_1 = \phi_1(Y^n)$$
$$\hat{U}_2 = \phi_2(Y^n, \hat{U}_1)$$
$$\hat{U}_3 = \phi_3(Y^n, \hat{U}^2)$$
$$\dots$$
$$\hat{U}_n = \phi_n(Y^n, \hat{U}^{n-1}).$$

If the genie-aided decoder makes no errors, then, the unaided decoder makes no errors.

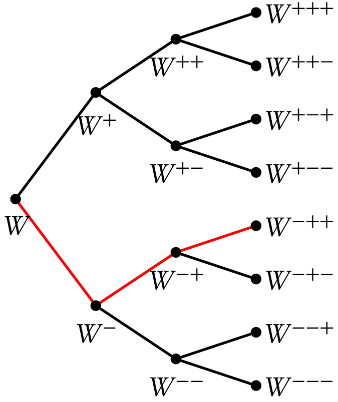
# Polarization does happen

- Let $B_1, B_2, \ldots$ be i.i.d., equally likely to be $\{+,-\}$, $W_0 = W$, $W_\ell = W_{\ell-1}^{B_\ell}$.



# Polarization does happen

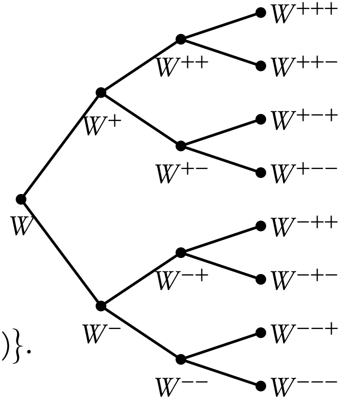
- Let $B_1, B_2, \ldots$ be i.i.d., equally likely to be $\{+, -\}$, $W_0 = W$, $W_\ell = W_{\ell-1}^{B_\ell}$.



# Polarization does happen

- Let $B_1, B_2, \dots$ be i.i.d., equally likely to be $\{+,-\}$, $W_0 = W$, $W_\ell = W_{\ell-1}^{B_\ell}$.
- $W_\ell$ is uniformly distributed among $\{W^{-\cdots-}, \dots, W^{+\cdots+}\}$, and

$$\Pr\big(I(W_\ell) \in (\epsilon, 1-\epsilon)\big) = \frac{1}{n}\#\big\{i \colon I(U_i; Y^n U^{i-1}) \in (\epsilon, 1-\epsilon)\big\}.$$



# Polarization does happen

To show polarization all we need to study is the process $I_\ell := I(W_\ell)$ and show that $I_\ell \in (\epsilon, 1-\epsilon)$ with small probability.

# Polarization does happen

To show polarization all we need to study is the process $I_\ell := I(W_\ell)$ and show that $I_\ell \in (\epsilon, 1-\epsilon)$ with small probability.

- $I_0 = I(W)$ is a constant.

# Polarization does happen

To show polarization all we need to study is the process $I_\ell := I(W_\ell)$ and show that $I_\ell \in (\epsilon, 1-\epsilon)$ with small probability.

- $I_0 = I(W)$ is a constant.
- $I_\ell$ lies in $[0, 1]$, so is bounded.

# Polarization does happen

To show polarization all we need to study is the process $I_\ell := I(W_\ell)$ and show that $I_\ell \in (\epsilon, 1-\epsilon)$ with small probability.

- $I_0 = I(W)$ is a constant.
- $I_\ell$ lies in $[0,1]$, so is bounded.
- Conditional on $B_1, \dots, B_\ell$, we know $W_\ell$, and $I_{\ell+1}$ is equally likely to be $I(W_\ell^-)$ and $I(W_\ell^+)$,

# Polarization does happen

To show polarization all we need to study is the process $I_\ell := I(W_\ell)$ and show that $I_\ell \in (\epsilon, 1-\epsilon)$ with small probability.

- $I_0 = I(W)$ is a constant.
- $I_\ell$ lies in $[0,1]$, so is bounded.
- Conditional on $B_1, \ldots, B_\ell$, we know $W_\ell$, and $I_{\ell+1}$ is equally likely to be $I(W_\ell^-)$ and $I(W_\ell^+)$,
- so,

$$E[I_{\ell+1}|B_1,\ldots,B_\ell] = \tfrac{1}{2}[I(W_\ell^-)+I(W_\ell^+)] = I(W_\ell) = I_\ell$$

and we see that $\{I_\ell\}$ is a martingale.

# Polarization does happen

- Bounded martingales converge almost surely.

# Polarization does happen

- Bounded martingales converge almost surely.
- Convergence of $\{I_\ell\}$ implies $|I_{\ell+1} - I_\ell| \to 0$.

# Polarization does happen

- Bounded martingales converge almost surely.
- Convergence of $\{I_\ell\}$ implies $|I_{\ell+1} - I_\ell| \to 0$.
- $|I_{\ell+1} - I_\ell| = \frac{1}{2}[I(W_\ell^+) - I(W_\ell^-)]$.

# Polarization does happen

- Bounded martingales converge almost surely.
- Convergence of $\{I_\ell\}$ implies $|I_{\ell+1} - I_\ell| \to 0$.
- $|I_{\ell+1} - I_\ell| = \frac{1}{2}[I(W_\ell^+) - I(W_\ell^-)]$.
- But $I(W_\ell^+) - I(W_\ell^-) < \delta$ implies $I(W_\ell) \notin (\epsilon, 1-\epsilon)$.

# Polarization does happen

- Bounded martingales converge almost surely.
- Convergence of $\{I_\ell\}$ implies $|I_{\ell+1} - I_\ell| \to 0$.
- $|I_{\ell+1} - I_\ell| = \frac{1}{2}[I(W_\ell^+) - I(W_\ell^-)]$.
- But $I(W_\ell^+) - I(W_\ell^-) < \delta$ implies $I(W_\ell) \notin (\epsilon, 1-\epsilon)$.
- Thus $I_\ell \to \{0, 1\}$, and

$$\Pr\big(I_\ell \in (\epsilon, 1-\epsilon)\big) \to 0.$$

# Polarization does happen

- Bounded martingales converge almost surely.
- Convergence of $\{I_\ell\}$ implies $|I_{\ell+1} - I_\ell| \to 0$.
- $|I_{\ell+1} - I_\ell| = \frac{1}{2}[I(W_\ell^+) - I(W_\ell^-)]$.
- But $I(W_\ell^+) - I(W_\ell^-) < \delta$ implies $I(W_\ell) \notin (\epsilon, 1-\epsilon)$.
- Thus $I_\ell \to \{0, 1\}$, and

$$\Pr\big(I_\ell \in (\epsilon, 1-\epsilon)\big) \to 0.$$

# Polarization speed

- We have seen that polarization takes place.

# Polarization speed

- We have seen that polarization takes place.
- But how fast? Fast enough to arrest error propagation?

# Polarization speed

- We have seen that polarization takes place.
- But how fast? Fast enough to arrest error propagation?
- Introduce the Bhattacharyya parameter

$$Z(W) = \sum_y \sqrt{W(y|0)W(y|1)}$$

as a companion to $I(W)$. Note that this is an upper bound on probability of error for uncoded transmission over $W$.

# Polarization speed

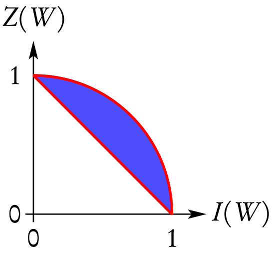
Properties of $Z(W)$:

- $Z(W) \in [0,1]$.

# Polarization speed

Properties of $Z(W)$:

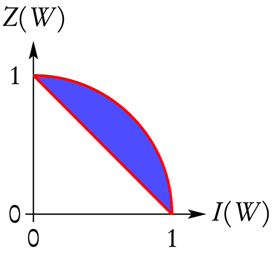
- $Z(W) \in [0, 1]$.
- $Z(W) \approx 0$ iff $I(W) \approx 1$.



# Polarization speed

Properties of $Z(W)$:

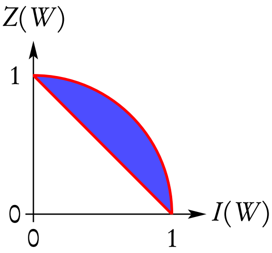
- $Z(W) \in [0, 1]$.
- $Z(W) \approx 0$ iff $I(W) \approx 1$.
- $Z(W) \approx 1$ iff $I(W) \approx 0$.



# Polarization speed

Properties of $Z(W)$:

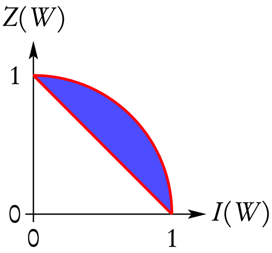
- $Z(W) \in [0, 1]$.
- $Z(W) \approx 0$ iff $I(W) \approx 1$.
- $Z(W) \approx 1$ iff $I(W) \approx 0$.
- $Z(W^+) = Z(W)^2$.



# Polarization speed

Properties of $Z(W)$:

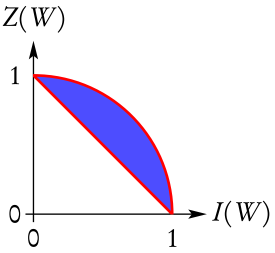
- $Z(W) \in [0, 1]$.
- $Z(W) \approx 0$ iff $I(W) \approx 1$.
- $Z(W) \approx 1$ iff $I(W) \approx 0$.
- $Z(W^+) = Z(W)^2$.
- $Z(W^-) \leq 2Z(W)$.



# Polarization speed

Properties of $Z(W)$:

- $Z(W) \in [0,1]$.
- $Z(W) \approx 0$ iff $I(W) \approx 1$.
- $Z(W) \approx 1$ iff $I(W) \approx 0$.
- $Z(W^+) = Z(W)^2$.
- $Z(W^-) \leq 2Z(W)$.




Since $Z(W)$ upper bounds on probability of error for uncoded transmission over $W$, we can choose the good indices on the basis of the $Z$'s of the synthetic channels. The sum of the $Z$'s of the chosen channels will upper bound the block error probability. This suggests studying the polarization speed of $Z$.

# Polarization speed

Given a binary input channel $W$,

# Polarization speed

Given a binary input channel $W$,

- just as for $I_\ell$, define $Z_\ell = Z(W^{B_1,\dots,B_\ell})$.

# Polarization speed

Given a binary input channel $W$,

- just as for $I_\ell$, define $Z_\ell = Z(W^{B_1,\dots,B_\ell})$.
- We know that $\Pr(Z_\ell \to 0) = I(W)$.

# Polarization speed

Given a binary input channel $W$,

- just as for $I_\ell$, define $Z_\ell = Z(W^{B_1,\ldots,B_\ell})$.
- We know that $\Pr(Z_\ell \to 0) = I(W)$.
- It turns out that when $Z_\ell \to 0$, it does so fast:

**Theorem**

*For any* $\beta < 1/2$, $\lim_{\ell\to\infty} \Pr(Z_\ell < 2^{-2^{\beta\ell}}) = I(W)$.

# Polarization speed

Given a binary input channel $W$,

- just as for $I_\ell$, define $Z_\ell = Z(W^{B_1,\ldots,B_\ell})$.
- We know that $\Pr(Z_\ell \to 0) = I(W)$.
- It turns out that when $Z_\ell \to 0$, it does so fast:

**Theorem**

*For any* $\beta < 1/2$, $\lim_{\ell\to\infty} \Pr(Z_\ell < 2^{-2^{\beta\ell}}) = I(W)$.

- This means that for any $\beta < 1/2$, as long as $R < I(W)$ the error probability of polarization codes decays to 0 faster than $2^{-n^\beta}$.

# So far

# So far

- Polar codes are $I(W)$ achieving,

# So far

- Polar codes are $I(W)$ achieving,
- encoding complexity is $n \log n$,

# So far

- Polar codes are $I(W)$ achieving,
- encoding complexity is $n \log n$,
- with successive decoding, the decoding complexity is $n \log n$,

# So far

- Polar codes are $I(W)$ achieving,
- encoding complexity is $n \log n$,
- with successive decoding, the decoding complexity is $n \log n$,
- probability of error decays like $2^{-\sqrt{n}}$.

# Moreover

# Moreover

- For symmetric channels the construction is deterministic. There is no "choose from this ensemble and verify" step.

# Moreover

- For symmetric channels the construction is deterministic. There is no "choose from this ensemble and verify" step.
- The error probability guarantees are not based on simulations — interesting for very low error probability applications.

# Moreover

- For symmetric channels the construction is deterministic. There is no “choose from this ensemble and verify” step.
- The error probability guarantees are not based on simulations — interesting for very low error probability applications.
- Generalizes to channels with arbitrary discrete input alphabets: a similar ‘two-by-two’ construction, with same complexity and error probability bounds. This allows one to achieve true capacity $C(W)$ rather than $I(W)$.

# Moreover

- For symmetric channels the construction is deterministic. There is no "choose from this ensemble and verify" step.
- The error probability guarantees are not based on simulations — interesting for very low error probability applications.
- Generalizes to channels with arbitrary discrete input alphabets: a similar 'two-by-two' construction, with same complexity and error probability bounds. This allows one to achieve true capacity $C(W)$ rather than $I(W)$.
- With '$k$-by-$k$' constructions the error probability can be made to decay almost exponentially in blocklength.

# Extensions

# Extensions

- Dual constructions yield vector quantizers that achieve the rate distortion bound.

# Extensions

- Dual constructions yield vector quantizers that achieve the rate distortion bound.
- Yields codes that achieve the secrecy capacity of the Wyner's wiretap channel.

# Extensions

- Dual constructions yield vector quantizers that achieve the rate distortion bound.
- Yields codes that achieve the secrecy capacity of the Wyner's wiretap channel.
- Polarizing each user of a MAC yield sum rate bound achieving codes.

# Extensions

- Dual constructions yield vector quantizers that achieve the rate distortion bound.
- Yields codes that achieve the secrecy capacity of the Wyner's wiretap channel.
- Polarizing each user of a MAC yield sum rate bound achieving codes.
- ....

# However

# However

- Exactly deciding which ones of the $n$ synthetic channels are good and which ones are bad is difficult.

# However

- Exactly deciding which ones of the $n$ synthetic channels are good and which ones are bad is difficult. But there are methods of low complexity that identify almost all the good channels and never misidentify a bad channel. (A vanishing fraction of good channels may be misidenfied as bad.)

# However

- Exactly deciding which ones of the $n$ synthetic channels are good and which ones are bad is difficult. But there are methods of low complexity that identify almost all the good channels and never misidentify a bad channel. (A vanishing fraction of good channels may be misidenfied as bad.)
- For asymptotic results to kick in one needs $\ell$ to be large ($n = 2^\ell$ is really large).

# However

- Exactly deciding which ones of the $n$ synthetic channels are good and which ones are bad is difficult. But there are methods of low complexity that identify almost all the good channels and never misidentify a bad channel. (A vanishing fraction of good channels may be misidenfied as bad.)
- For asymptotic results to kick in one needs $\ell$ to be large ($n = 2^\ell$ is really large).
- For channel coding applications, the usual suspects (LDPC, Turbo, ...) easily beat the pure polar codes at meaningful block lengths.

# However

- Exactly deciding which ones of the $n$ synthetic channels are good and which ones are bad is difficult. But there are methods of low complexity that identify almost all the good channels and never misidentify a bad channel. (A vanishing fraction of good channels may be misidenfied as bad.)
- For asymptotic results to kick in one needs $\ell$ to be large ($n = 2^\ell$ is really large).
- For channel coding applications, the usual suspects (LDPC, Turbo, ...) easily beat the pure polar codes at meaningful block lengths. For quantization they are a lot more competitive.

# Remarks

- Polar codes are close cousins of Reed–Muller codes. Differ only in the choice of which indices are good.

## Remarks

- Polar codes are close cousins of Reed–Muller codes. Differ only in the choice of which indices are good.
- One transform $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}^{\otimes \ell}$ works for all channels.

# Remarks

- Polar codes are close cousins of Reed–Muller codes. Differ only in the choice of which indices are good.
- One transform $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}^{\otimes \ell}$ works for all channels.
- The technique treats noise not by eliminating it, but by shifting it to a subspace.

# Remarks

- Polar codes are close cousins of Reed–Muller codes. Differ only in the choice of which indices are good.
- One transform $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}^{\otimes \ell}$ works for all channels.
- The technique treats noise not by eliminating it, but by shifting it to a subspace.
- Successive decoding is cheap ($n \log n$) but too naive. More clever decoding methods should improve error probability at moderate $n$.

# Remarks

- Polar codes are close cousins of Reed–Muller codes. Differ only in the choice of which indices are good.
- One transform $\left[\begin{smallmatrix}1 & 1\\ 0 & 1\end{smallmatrix}\right]^{\otimes \ell}$ works for all channels.
- The technique treats noise not by eliminating it, but by shifting it to a subspace.
- Successive decoding is cheap ($n \log n$) but too naive. More clever decoding methods should improve error probability at moderate $n$.
- While the original motivation was for channel coding, polar codes are good at many settings. [“Polar codes are good for everything — S. Korada”]