Fast Polarization for Processes with Memory

Boaz Shuval, Ido Tal
Technion

Polar codes for busy people

$X_1^N \longrightarrow \boxed{W} \longrightarrow Y_1^N$
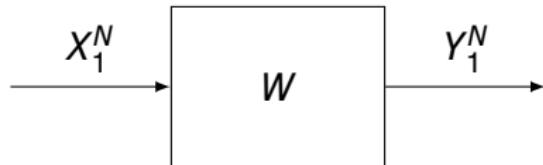
Polar coding

- ▶ **Information vector**: $\tilde{U}_1^k$
- ▶ **Padding**: $U_1^N = f(\tilde{U}_1^k)$
- ▶ **Encoding**: $X_1^N = U_1^N \cdot G_N^{-1}$
- ▶ **Decoding**: Successively, deduce U_i from U_1^{i-1} and Y_1^N

Polar codes: [Arıkan:09], [ArıkanTelatar:09]

- **Setting**: binary-input, symmetric, memoryless channel
- **Polar transform**: $U_1^N = X_1^N \cdot G_N$

$$X_1^N \quad \text{uniform} \quad \Longleftrightarrow \quad U_1^N \quad \text{uniform}$$

- **Low entropy indices**: Fix $\beta < 1/2$

$$\Lambda_N = \left\{ i : P_{\text{error}}(U_i | U_1^{i-1}, Y_1^N) < 2^{-N^\beta} \right\}$$

- **Polarization**: Let X_1^N be uniform

$$\lim_{N \to \infty} \frac{1}{N} |\Lambda_N| = I(W)$$

- **Coding scheme**:
  - For $i \in \Lambda_N$, set U_i equal to information bits (uniform)
  - Set remaining U_i to uniform values, reveal to decoder
  - Transmit $X_1^N = U_1^N \cdot G_N^{-1}$ as codeword

In this talk

Setting: binary-input, symmetric, memoryless channel

In this talk

Setting: ~~binary-input~~, symmetric, memoryless channel

In this talk

Setting: ~~binary-input~~, ~~symmetric~~, memoryless channel

In this talk

Setting: ~~binary-input~~, ~~symmetric~~, memoryless ~~channel~~

In this talk

Setting: ~~binary-input~~, ~~symmetric~~, ~~memoryless~~ channel

Polar codes: [Şaşoğlu+:09], [KoradaUrbanke:10], [HondaYamamoto:13]

- **Setting**: Memoryless i.i.d. process $(X_i, Y_i)_{i=1}^N$
- **For simplicity**: Assume X_i binary
- **Polar transform**: $U_1^N = X_1^N \cdot G_N$
- **Index sets**:

  Low entropy: $\quad \Lambda_N = \left\{ i : P_{\text{error}}(U_i | U_1^{i-1}, Y_1^N) < 2^{-N^\beta} \right\}$

  High entropy: $\quad \Omega_N = \left\{ i : P_{\text{error}}(U_i | U_1^{i-1}, Y_1^N) > \frac{1}{2} - 2^{-N^\beta} \right\}$

- **Polarization**:

$$\lim_{N \to \infty} \frac{1}{N} |\Lambda_N| = 1 - H(X_1 | Y_1)$$
$$\lim_{N \to \infty} \frac{1}{N} |\Omega_N| = H(X_1 | Y_1)$$

Polar codes: [Şaşoğlu+:09], [KoradaUrbanke:10], [HondaYamamoto:13]

Optimal rate for:

- ▶ Coding for non-symmetric memoryless channels
- ▶ Coding for memoryless channels with non-binary inputs
- ▶ (Lossy) compression of memoryless sources

Question

- ▶ How to handle memory?

A framework for memory

- **Process**:

$$(X_i, Y_i, S_i)_{i=1}^{N}$$

  - **Finite number of states**: $S_i \in \mathcal{S}$, where $|\mathcal{S}| < \infty$
  - **Hidden state**: S_i is unknown to encoder and decoder

- **Probability distribution**:

$$P(x_i, y_i, s_i | s_{i-1})$$

  - **Stationary**: same for all i
  - **Markov**:

$$P(x_i, y_i, s_i | s_{i-1}) = P(x_i, y_i, s_i | \{x_j, y_j, s_j\}_{j<i})$$

  - **State sequence**: aperiodic and irreducible Markov chain

Example 1

- **Model**: Finite state channel

$$P_s(y|x) , \quad s \in \mathcal{S}$$

- **Input distribution**: X_i i.i.d. and independent of state
- **State transition**:

$$\pi(s_i|s_{i-1})$$

- **Distribution**:

$$P(x_i, y_i, s_i|s_{i-1}) = P(x_i)\pi(s_i|s_{i-1})P_{s_i}(y_i|x_i)$$

Example 2

- **Model**: ISI + noise

$$Y_i = h_0 X_i + h_1 X_{i-1} + \cdots + h_m X_{i-m} + \text{noise}$$

- **Input**: X_i has memory

$$P(x_i | x_{i-1}, x_{i-2}, \ldots, x_{i-m}, x_{i-m-1})$$

- **State**:

$$S_i = \begin{bmatrix} X_i & X_{i-1} & \cdots & X_{i-m} \end{bmatrix}$$

- **Distribution**: For x_i, s_i, s_{i-1} compatible,

$$P(x_i, y_i, s_i | s_{i-1}) = P_{\text{noise}}(y_i | h^T s_i) \cdot P(x_i | s_{i-1})$$

Example 3

- **Model**: (d, k)-RLL constrained system with noise
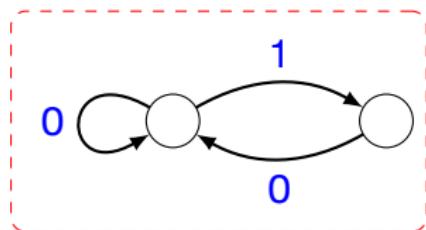
$(1, \infty)$-RLL Constraint

Example 3

- **Model**: (d, k)-RLL constrained system with noise


$(1, \infty)$-RLL Constraint

Example 3

▶ **Model**: (d, k)-RLL constrained system with noise

$(1, \infty)$-RLL Constraint

state Markov chain

$P(x_i, y_i, s_i | s_{i-1})$
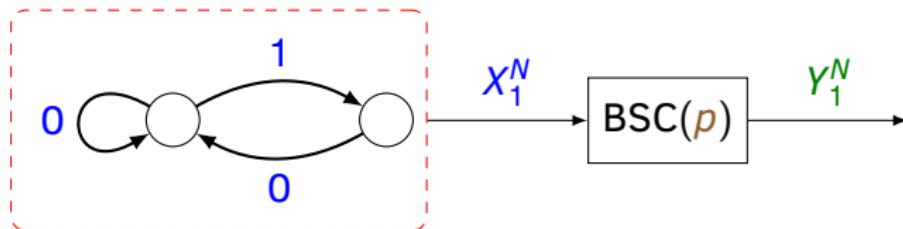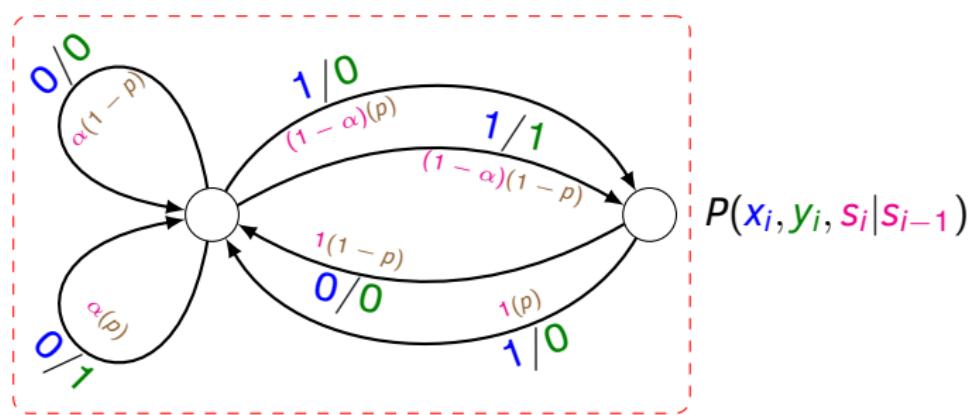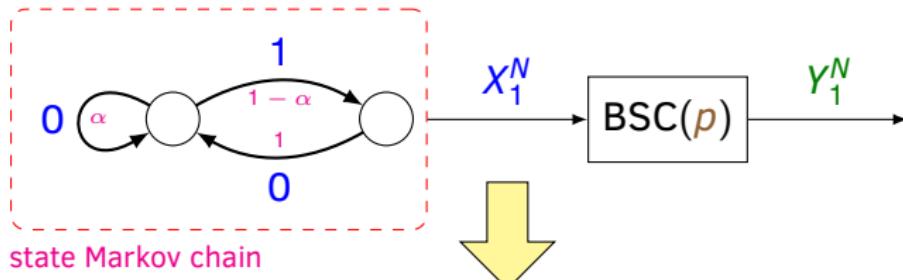
Example 4

- **Model**: Lossy compression of a source with memory
- **Input distribution**:

$$P_{s_i}(x_i) \,, \quad s_i \in \mathcal{S}$$

- **State transition**:

$$\pi(s_i|s_{i-1})$$

- **Distortion**: test channel ($Y_i \equiv \hat{X}_i$)

$$P(y_i|x_i)$$

- **Distribution**:

$$P(x_i, y_i, s_i|s_{i-1}) = \pi(s_i|s_{i-1})P_{s_i}(x_i)P(y_i|x_i)$$

Polar codes: [Şaşoğlu:11], [ŞaşoğluTal:16], [ShuvalTal:17]

- **Setting**: Process $(X_i, Y_i, S_i)_{i=1}^N$ with memory, as above
- **Hidden state**: State unknown to encoder and decoder
- **Polar transform**: $U_1^N = X_1^N \cdot G_N$
  U_1^N are neither independent, nor identically distributed
- **Index sets**:

  Low entropy: $\quad \Lambda_N = \left\{ i : P_{\text{error}}(U_i | U_1^{i-1}, Y_1^N) < 2^{-N^\beta} \right\}$

  High entropy: $\quad \Omega_N = \left\{ i : P_{\text{error}}(U_i | U_1^{i-1}, Y_1^N) > \frac{1}{2} - 2^{-N^\beta} \right\}$

- **Polarization**:

$$\lim_{N \to \infty} \frac{1}{N} |\Lambda_N| = 1 - H_\star(X|Y)$$

$$\lim_{N \to \infty} \frac{1}{N} |\Omega_N| = H_\star(X|Y)$$

$$\boxed{\lim_{N \to \infty} \frac{1}{N} H(X_1^N | Y_1^N)}$$

Achievable rate

▸ **Achievable rate**: In all examples, R approaches

$$I_\star(X; Y) = \lim_{N \to \infty} \frac{1}{N} I(X_1^N; Y_1^N)$$

▸ **Successive cancellation**: [Wang+:15]

▸ **Not needed**: ~~State estimation~~
Don't waste precious bits on this!

Three parameters

- Joint distribution $P(x, y)$
- For simplicity: $X \in \{0, 1\}$
- **Parameters**:

  Entropy $\qquad H(X|Y) = -\sum_{x,y} P(x,y) \log P(x|y)$

  Bhattacharyya $\qquad Z(X|Y) = 2 \sum_y \sqrt{P(0,y)P(1,y)}$

  T.V. distance $\qquad K(X|Y) = \sum_y |P(0,y) - P(1,y)|$

- **Connections**:

$$H \approx 0 \iff Z \approx 0 \iff K \approx 1$$

$$H \approx 1 \iff Z \approx 1 \iff K \approx 0$$

Three processes

For $n = 1, 2, \ldots$

- $N = 2^n$
- $U_1^N = X_1^N G_N$
- Pick $B_n \in \{0, 1\}$ uniform, i.i.d.
- **Random index** from $\{1, 2, \ldots, N\}$
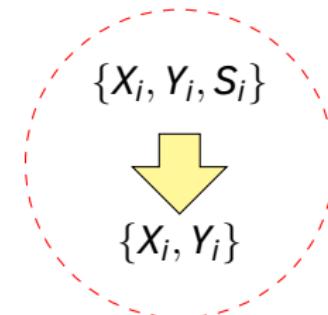
$$i = 1 + \langle B_1 \, B_2 \, \cdots \, B_n \rangle_2$$

- **Processes**:

Entropy	$H_n = H(U_i | U_1^{i-1}, Y_1^N)$
Bhattacharyya	$Z_n = Z(U_i | U_1^{i-1}, Y_1^N)$
T.V. distance	$K_n = K(U_i | U_1^{i-1}, Y_1^N)$

$\{X_i, Y_i, S_i\}$

$\Downarrow$

$\{X_i, Y_i\}$

Proof — memoryless case

Slow polarization ⟶ Fast polarization

$$H_n \in (\epsilon, 1 - \epsilon)$$

$$\Downarrow$$
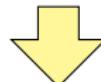
$$|H_{n+1} - H_n| > 0$$

$$Z_{n+1} \leq \begin{cases} 2Z_n & B_{n+1} = 0 \\ Z_n^2 & B_{n+1} = 1 \end{cases}$$

$$\Downarrow$$

$$\frac{1}{N}|\Lambda_N| \xrightarrow[n\to\infty]{} 1 - H(X_1|Y_1)$$

$$K_{n+1} \leq \begin{cases} K_n^2 & B_{n+1} = 0 \\ 2K_n & B_{n+1} = 1 \end{cases}$$

$$\Downarrow$$

$$\frac{1}{N}|\Omega_N| \xrightarrow[n\to\infty]{} H(X_1|Y_1)$$

Proof — memoryless case

Slow polarization   Fast polarization

$H_n \in (\epsilon, 1 - \epsilon)$

$\Downarrow$

$|H_{n+1} - H_n| > 0$

$\psi = \max_s \dfrac{1}{\pi(s)}$
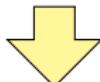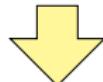
π: stationary
state distribution

$Z_{n+1} \leq \begin{cases} 2\psi Z_n & B_{n+1} = 0 \\ \psi Z_n^2 & B_{n+1} = 1 \end{cases}$

$\Downarrow$

$\dfrac{1}{N}|\Lambda_N| \xrightarrow[n\to\infty]{} 1 - H_\star(X|Y)$
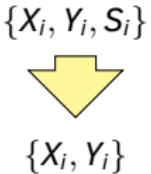
$\hat{K}_{n+1} \leq \begin{cases} \psi \hat{K}_n^2 & B_{n+1} = 0 \\ 2\hat{K}_n & B_{n+1} = 1 \end{cases}$

$\Downarrow$
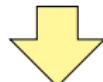
$\dfrac{1}{N}|\Omega_N| \xrightarrow[n\to\infty]{} H_\star(X|Y)$

Unmemorizing

- **Memoryless case**:
  - Proof hinges on independence:

  $$P(x_1^{2N}, y_1^{2N}) = P(x_1^N, y_1^N) \cdot P(x_{N+1}^{2N}, y_{N+1}^{2N})$$

- **Memory case**:
 - Force independence: condition on middle state S_N

  $$P(x_1^{2N}, y_1^{2N}|s_N) = P(x_1^N, y_1^N|s_N) \cdot P(x_{N+1}^{2N}, y_{N+1}^{2N}|s_N)$$

- **New processes**:

 $$\hat{H}_n = H(U_i|U_1^{i-1}, Y_1^N, S_0, S_N)$$
  $$\hat{K}_n = K(U_i|U_1^{i-1}, Y_1^N, S_0, S_N)$$

Tying things together

$$Z_n \longleftarrow H_n$$

$$\downarrow$$

$$\Lambda_N$$
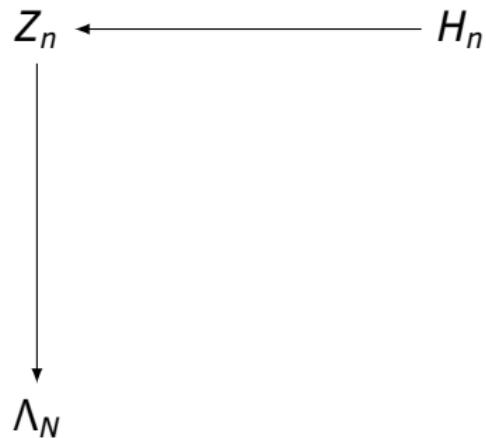
Tying things together
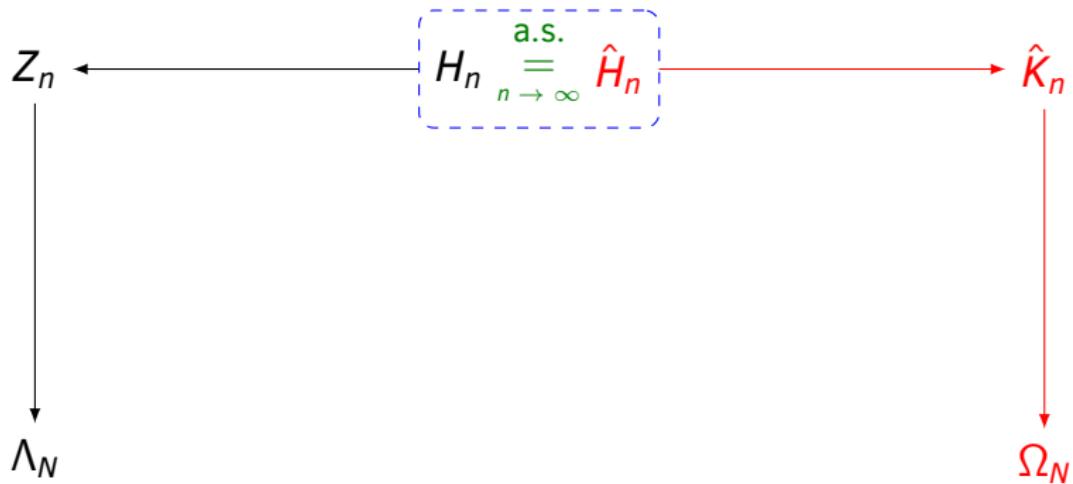
Polarization of K_n (memoryless case)

- **Two consecutive blocks**: (X_1^N, Y_1^N) and $(X_{N+1}^{2N}, Y_{N+1}^{2N})$
- **Polar transform**:

$$U_1^N = X_1^N \cdot G_N$$
$$V_1^N = X_{N+1}^{2N} \cdot G_N$$

- **Random index**:

$$i = 1 + \langle B_1 \, B_2 \, \cdots \, B_n \rangle_2$$

- **Notation**:

$$Q_i = (U_1^{i-1}, Y_1^N)$$
$$R_i = (V_1^{i-1}, Y_{N+1}^{2N})$$

Polarization of K_n (memoryless case)

- **Memoryless** assumption:

$$P(u_i, v_i, q_i, r_i) = P(u_i, q_i) \cdot P(v_i, r_i)$$

- **Notation**:

$$T_i = U_i + V_i$$

- **One step polarization**:

$$K_{n+1} = \begin{cases} K(T_i | Q_i, R_i) & B_{n+1} = 0 & \text{`$-$' transform} \\ K(V_i | T_i, Q_i, R_i) & B_{n+1} = 1 & \text{`$+$' transform} \end{cases}$$

- **Recall**:

$$K(X|Y) = \sum_y |P(0, y) - P(1, y)|$$

Polarization of K_n (memoryless case), '$-$' transform

$$K_{n+1} = \sum_{q,r} |P_{\mathsf{T}_i,\mathsf{Q}_i,\mathsf{R}_i}(0,q,r) - P_{\mathsf{T}_i,\mathsf{Q}_i,\mathsf{R}_i}(1,q,r)|$$

$$= \sum_{q,r} \left| \sum_{v=0}^{1} P(v,r)(P(v,q) - P(v+1,q)) \right|$$

$$= \sum_{q,r} \left| \Big(P(0,q) - P(1,q)\Big)\Big(P(0,r) - P(1,r)\Big) \right|$$

$$= \sum_{q,r} |P(0,q) - P(1,q)| \cdot |P(0,r) - P(1,r)|$$

$$= \sum_{q} |P(0,q) - P(1,q)| \cdot \sum_{r} |P(0,r) - P(1,r)|$$

$$= K_n^2,$$

Polarization of K_n (memoryless case), '+' transform

$$K_{n+1} = \sum_{t,q,r} |P_{\mathrm{T}_i,\mathrm{V}_i,\mathrm{Q}_i,\mathrm{R}_i}(t,0,q,r) - P_{\mathrm{T}_i,\mathrm{V}_i,\mathrm{Q}_i,\mathrm{R}_i}(t,1,q,r)|$$

$$= \sum_{t,q,r} |P(t,q)P(0,r) - P(t+1,q)P(1,r)|$$

$$\overset{(*)}{\leq} \frac{1}{2} \sum_{t,q,r} P(q)\,|P(0,r) - P(1,r)| + P(r)\,|P(t,q) - P(t+1,q)|$$

$$= \frac{1}{2} \sum_{t,r} |P(0,r) - P(1,r)| + \frac{1}{2} \sum_{t,q} |P(t,q) - P(t+1,q)|$$

$$= 2K_n,$$

Identity for $(*)$: For any a,b,c,d:

$$ab - cd = \frac{(a+c)(b-d) + (b+d)(a-c)}{2}$$

Polarization of $\hat{K}_n$ (memory)

- ▶ Follows steps of memoryless case
- ▶ Requires additional inequalities
  - ▶ **Inequality I**: For states $s_0, s_N, s_{2N} \in \mathcal{S}$,

  $$
  P(s_0, s_N, s_{2N}) = \frac{P(s_0, s_N) \cdot P(s_N, s_{2N})}{P(s_N)}
  $$
  $$
  \leq \psi \cdot P(s_0, s_N) \cdot P(s_N, s_{2N})
  $$

  where

  $$
  \psi = \max_s \frac{1}{\pi(s)}
  $$

  - ▶ **Inequality II**: For $f, g \geq 0$,

  $$
  \sum_{s_N} f(s_N) g(s_N) \leq \sum_{s_N} f(s_N) \sum_{s_N'} g(s_N')
  $$

Summary

- A general framework for memory:

$$P(x_i, y_i, s_i | s_{i-1})$$

  - Memory allowed in both source and channel
  - State sequence S_i
    - Hidden
    - Stationary
    - Finite state Markov
    - Aperiodic and irreducible
- Achieve rate $I_\star(X; Y)$ through polar codes
- No change to polarization exponent ($\beta < 1/2$)

Applicable to many real-life communication scenarios