Universal Polarization for Processes with Memory

Boaz Shuval and Ido Tal

Andrew and Erna Viterbi Department of Electrical Engineering
Technion — Israel Institute of Technology
Haifa, 32000, Israel

July 2019

Setting

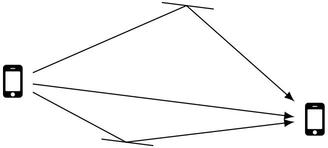
- Communication with uncertainty:
  - Encoder: Knows channel belongs to a set of channels
  - Decoder: Knows channel statistics (e.g., via estimation)
- Memory:
  - In channels
  - In input distribution
- Universal code:
  - Vanishing error probability over set
  - Best rate (infimal information rate over set)

Goal:

Universal Code based on Polarization

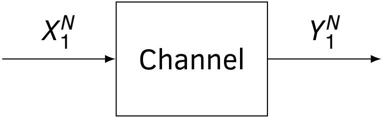
Why?

- Polar codes have many good properties
  - rate-optimal (even under memory!)
  - vanishing error probability
  - low complexity encoding/decoding/construction
- But...
  - Polar codes must be tailored to the channel at hand
- Sometimes, the channel isn't known apriori to encoder
  - **Example:** Frequency Selective Fading $\Rightarrow$ ISI



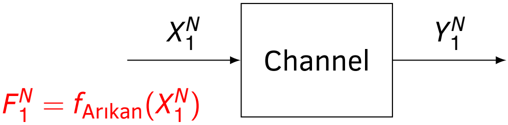
$$Y_n = h_0 X_n + \sum_{i=1}^{m} h_i X_{n-i} + \text{noise}$$

Polar Codes: lightning reminder



- **Goal:** Decode X_1^N from Y_1^N

Polar Codes: lightning reminder



- **Goal:** Decode X_1^N from Y_1^N
- **Transform** $f_{\text{Arıkan}}$ is one-to-one and onto
  - recursively defined
- Decoding $X_1^N \Longleftrightarrow$ Decoding F_1^N

Polar Codes: lightning reminder

- **Successive-Cancellation decoding:**
  - Compute G_i from decoded F_1^{i-1}
  - Decode F_i from G_i
- **Polarization:** fix $\beta < 1/2$
  - Low-Entropy set: $\mathcal{L}_N = \{i \mid H(F_i|G_i) < 2^{-N^\beta}\}$
  - High-Entropy set: $\mathcal{H}_N = \{i \mid H(F_i|G_i) > 1 - 2^{-N^\beta}\}$
  - For N large, $|\mathcal{L}_N| + |\mathcal{H}_N| \approx N$
- **Coding scheme** (simplified):
  - $i \in \mathcal{L}_N \Rightarrow$ Transmit data
  - $i \in \mathcal{H}_N \Rightarrow$ Reveal to decoder

Polar Codes: lightning reminder

- **Successive-Cancellation decoding:**
  - Compute G_i from decoded F_1^{i-1}
  - Decode F_i from G_i
- **Polarization:** fix $\beta < 1/2$
  - Low-Entropy set: $\mathcal{L}_N = \{i \mid H(F_i|G_i) < 2^{-N^\beta}\}$
  - High-Entropy set: $\mathcal{H}_N = \{i \mid H(F_i|G_i) > 1 - 2^{-N^\beta}\}$
  - For N large, $|\mathcal{L}_N| + |\mathcal{H}_N| \approx N$
- **Coding scheme** (simplified):
  - $i \in \mathcal{L}_N \Rightarrow$ Transmit data
  - $i \in \mathcal{H}_N \Rightarrow$ Reveal to decoder

Not Universal!
$\mathcal{L}_N, \mathcal{H}_N$ channel-dependent

Previous Work on Universal Polarization

- All for the memoryless case
- Works with **memoryless** settings similar to ours:
  - Hassani & Urbanke 2014
  - Şaşoğlu& Wang 2016 (conference version: 2014)

Previous Work on Universal Polarization

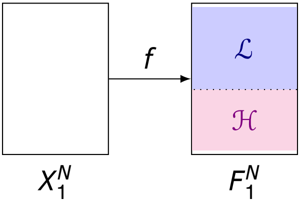
- All for the memoryless case
- Works with **memoryless** settings similar to ours:
  - Hassani & Urbanke 2014
  - Şaşoğlu& Wang 2016 (conference version: 2014)

Our Construction

- Simplified generalization of Şaşoğlu-Wang construction
- Memory at channel and/or input
- Two stages: “slow” and “fast”

Our Construction

- Simplified generalization of Şaşoğlu-Wang construction
- Memory at channel and/or input
- Two stages: "slow" and "fast"

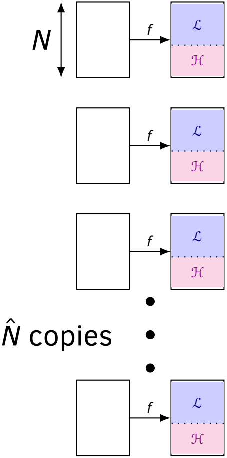


- f one-to-one and onto, recursively defined
- $(\eta, \mathcal{L}, \mathcal{H})$-monopolarization:
  For any $\eta > 0$, there exist N and index sets $\mathcal{L}, \mathcal{H}$ such that

  either $\quad H(F_i|G_i) < \eta \quad$ for all $i \in \mathcal{L}$
  or $\quad H(F_i|G_i) > 1 - \eta \quad$ for all $i \in \mathcal{H}$
- **Universal**: $\mathcal{L}$, $\mathcal{H}$ process independent
- **Slow**

Our Construction

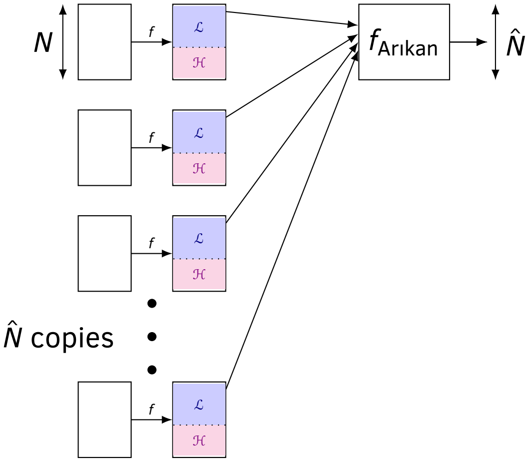
- Simplified generalization of Şaşoğlu-Wang construction
- Memory at channel and/or input
- Two stages: "slow" and "fast"



$i \in \mathcal{L} \Rightarrow H(F_i|G_i) < \eta$

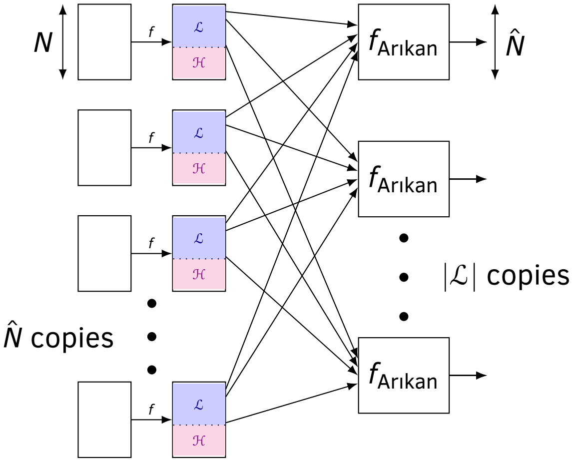
Our Construction

- Simplified generalization of Şaşoğlu-Wang construction
- Memory at channel and/or input
- Two stages: "slow" and "fast"



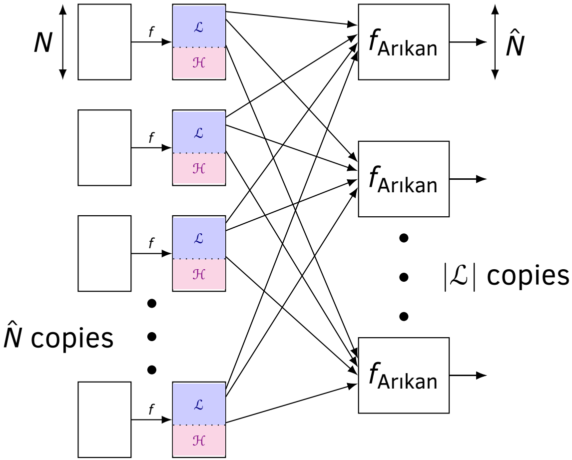
Our Construction

- Simplified generalization of Şaşoğlu-Wang construction
- Memory at channel and/or input
- Two stages: "slow" and "fast"



Our Construction

- Simplified generalization of Şaşoğlu-Wang construction
- Memory at channel and/or input
- Two stages: "slow" and "fast"



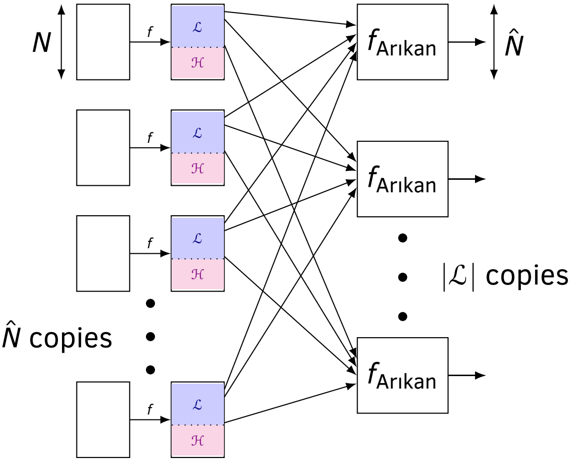
$$P_e \leq |\mathcal{L}| \cdot 2^{-\hat{N}^{\beta}}$$

$$\text{Rate} \approx \frac{|\mathcal{L}|}{N}$$

Our Construction

- Simplified generalization of Şaşoğlu-Wang construction
- Memory at channel and/or input
- Two stages: "slow" and "fast"

Our focus



$$P_e \leq |\mathcal{L}| \cdot 2^{-\hat{N}^\beta}$$

$$\text{Rate} \approx \frac{|\mathcal{L}|}{N}$$

A framework for memory

- Stationary process:

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

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

- Markov property:

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

- FAIM state sequence:
  Finite-state, **a**periodic, **i**rreducible **M**arkov chain
- $(X_i, Y_i)_{i=1}^N$ FAIM-derived process
- FAIM $\Rightarrow$ mixing: if $M - N$ large enough,
  $(X_{-\infty}^N, Y_{-\infty}^N)$ and (X_M^∞, Y_M^∞) almost independent

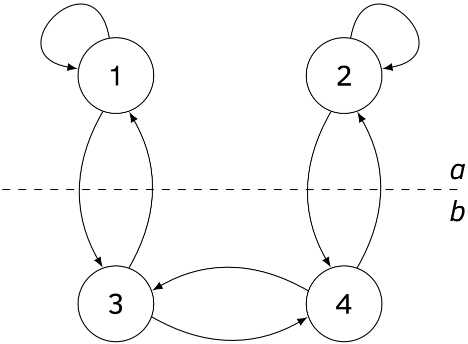
Forgetfulness

- Required for proof of monopolarization
- FAIM process (S_i, X_i, Y_i) is forgetful if for any $\epsilon > 0$ there exists natural λ such that if $k \geq \lambda$,

$$I(S_1; S_k | X_1^k, Y_1^k) \leq \epsilon$$
$$I(S_1; S_k | Y_1^k) \leq \epsilon$$

- Neither inequality implies the other
- FAIM does not imply forgetfulness
- We have a sufficient condition for forgetfulness
  - Under it, ϵ decreases exponentially with λ

FAIM Does Not Imply Forgetfulness



$$Y_j = \begin{cases} a, & S_j \in \{1, 2\} \\ b, & S_j \in \{3, 4\} \end{cases}$$

$$I(S_1; S_k | Y_1^k) \not\to 0$$

Why Forgetfulness?

- (S_i, X_i, Y_i) forgetful if for any $\epsilon > 0$ exists λ such that

$$k \geq \lambda \implies \begin{cases} I(S_1; S_k | X_1^k, Y_1^k) \leq \epsilon \\ I(S_1; S_k | Y_1^k) \leq \epsilon \end{cases}$$

- Can show: for any $k + 1 \leq i \leq N - k$

$$0 \leq H(X_i | X_{i-k}^{i-1}, Y_{i-k}^{i+k}) - H(X_i | X_1^{i-1}, Y_1^N) \leq 2\epsilon$$

Takeaway point

Only a “window” surrounding i really matters

Slow Stage is Monopolarizing

- FAIM-derived: (X_i, Y_i) derived from (S_i, X_i, Y_i) such that

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

  with S_i finite-state, aperiodic, irreducible, Markov
- Forgetful: for any $\epsilon > 0$ there exists λ such that if $k \geq \lambda$,

$$I(S_1; S_k | X_1^k, Y_1^k) \leq \epsilon$$
$$I(S_1; S_k | Y_1^k) \leq \epsilon$$

Main Result (simplified)

If process (X_i, Y_i) is FAIM-derived and forgetful, the slow stage is monopolarizing, with universal $\mathcal{L}, \mathcal{H}$ (unrelated to process)

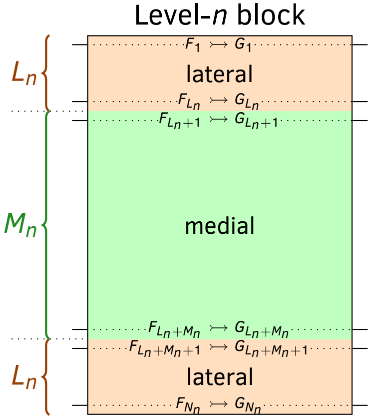
Slow Stage

- Presented for the case $|\mathcal{L}| = |\mathcal{H}|$
- Transforms

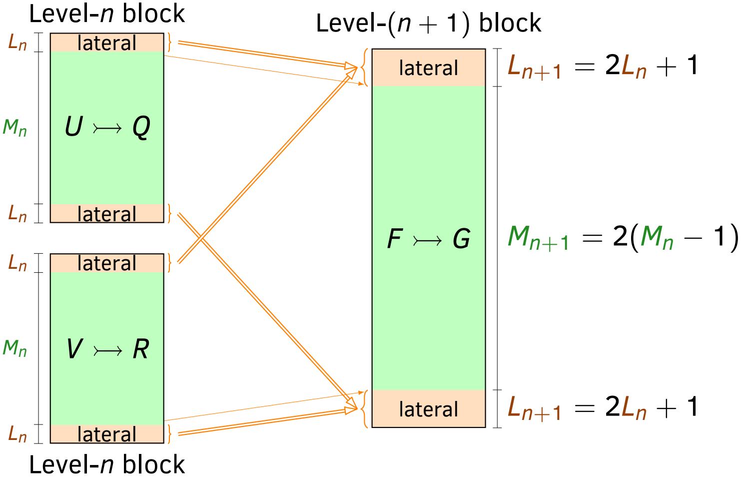
$$X_1^{N_n} \rightarrowtail Y_1^{N_n} \overset{f}{\Rightarrow} F_1^{N_n} \rightarrowtail G_1^{N_n}$$

transmitted received decode F_i from G_i

- Recursively defined
  - Parameters L_0, M_0
  - Level 0 length: $N_0 = 2L_0 + M_0$
  - Level n length: $N_n = 2N_{n-1}$
- Index types at level n:
  - First L_n indices: lateral
  - Middle M_n indices: medial
  - Last L_n indices: lateral



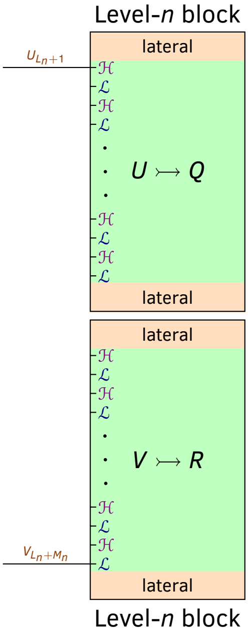
Slow Stage — Lateral Recursion



- Lateral indices always remain lateral
- Two medial indices become lateral

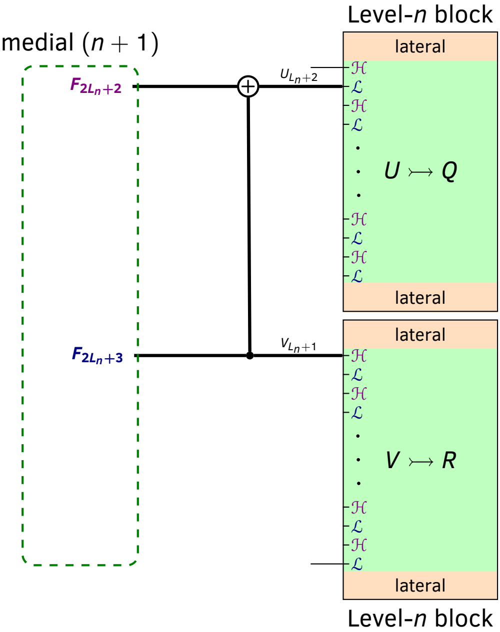
Slow Stage — Medial Recursion

- Two type of medial indices:
  - $\mathcal{H}$
  - $\mathcal{L}$
- Alternating:
  $\mathcal{H}, \mathcal{L}, \mathcal{H}, \mathcal{L}, \ldots$
- Two medial become lateral:
  $U_{L_n+1}, V_{L_n+M_n}$
- Join $\mathcal{H}$ from one block with $\mathcal{L}$ from other



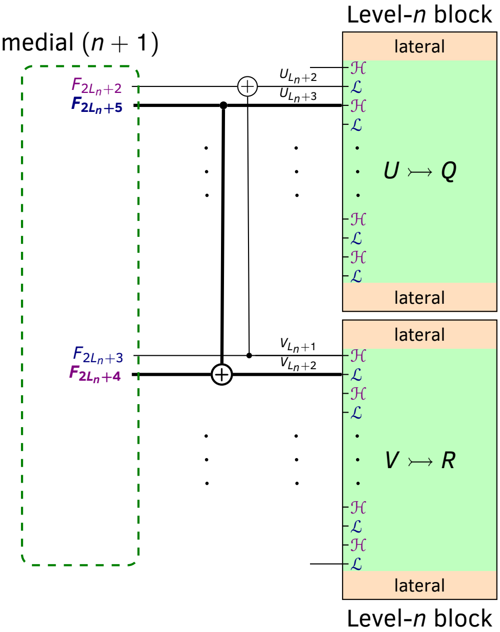
Slow Stage — Medial Recursion

- Two type of medial indices:
  - $\mathcal{H}$
  - $\mathcal{L}$
- Alternating:
  $\mathcal{H}, \mathcal{L}, \mathcal{H}, \mathcal{L}, \ldots$
- Two medial become lateral:
  $U_{L_n+1}, V_{L_n+M_n}$
- Join $\mathcal{H}$ from one block with $\mathcal{L}$ from other



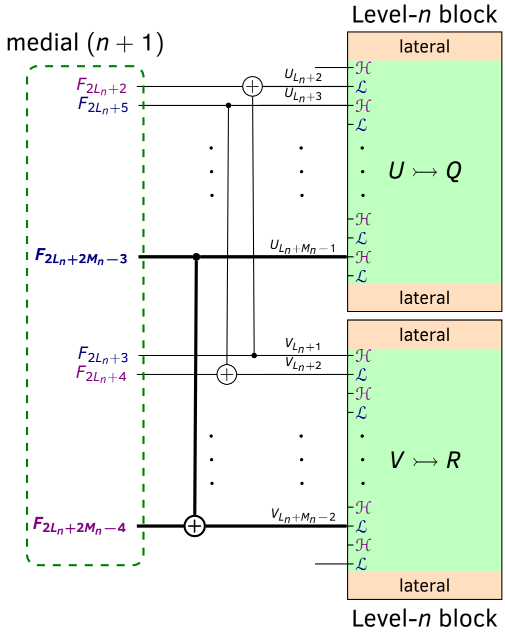
Slow Stage — Medial Recursion

- Two type of medial indices:
  - $\mathcal{H}$
  - $\mathcal{L}$
- Alternating: $\mathcal{H}, \mathcal{L}, \mathcal{H}, \mathcal{L}, \ldots$
- Two medial become lateral: $U_{L_n+1}, V_{L_n+M_n}$
- Join $\mathcal{H}$ from one block with $\mathcal{L}$ from other



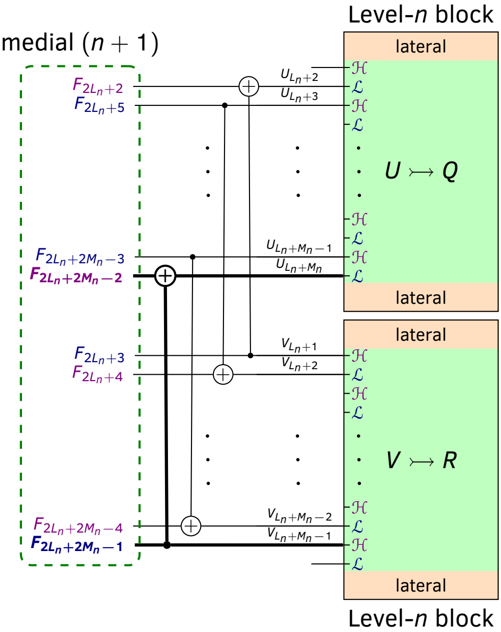
Slow Stage — Medial Recursion

- Two type of medial indices:
  - $\mathcal{H}$
  - $\mathcal{L}$
- Alternating: $\mathcal{H}, \mathcal{L}, \mathcal{H}, \mathcal{L}, \ldots$
- Two medial become lateral: $U_{L_n+1}, V_{L_n+M_n}$
- Join $\mathcal{H}$ from one block with $\mathcal{L}$ from other



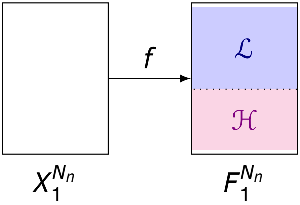
Slow Stage — Medial Recursion

- Two type of medial indices:
  - $\mathcal{H}$
  - $\mathcal{L}$
- Alternating: $\mathcal{H}, \mathcal{L}, \mathcal{H}, \mathcal{L}, \ldots$
- Two medial become lateral: $U_{L_n+1}, V_{L_n+M_n}$
- Join $\mathcal{H}$ from one block with $\mathcal{L}$ from other



Slow Stage is Monopolarizing

Slow stage:



Main Result

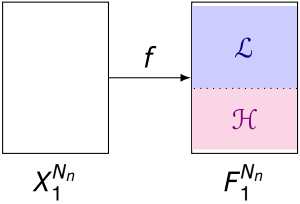
If process (X_i, Y_i) is FAIM-derived and forgetful, for every $\eta > 0$, there exist L_0, M_0, n_{th} such that the slow stage of level at least n_{th} is $(\eta,\mathcal{L},\mathcal{H})$-monopolarizing

$$H_\star(X|Y) \leq 1/2 \Rightarrow H(F_i|G_i) < \eta \quad \text{for all } i \in \mathcal{L}$$

$$H_\star(X|Y) \geq 1/2 \Rightarrow H(F_i|G_i) > 1 - \eta \quad \text{for all } i \in \mathcal{H}$$

Slow Stage is Monopolarizing

Slow stage:



Main Result

If process (X_i, Y_i) is FAIM-derived and forgetful, for every $\eta > 0$, there exist L_0, M_0, n_{th} such that the slow stage of level at least n_{th} is $(\eta,\mathcal{L},\mathcal{H})$-monopolarizing

$$H_\star(X|Y) \leq 1/2 \Rightarrow H(F_i|G_i) < \eta \qquad \text{for all } i \in \mathcal{L}$$
$$H_\star(X|Y) \geq 1/2 \Rightarrow H(F_i|G_i) > 1 - \eta \qquad \text{for all } i \in \mathcal{H}$$

Universal: sets $\mathcal{L}$, $\mathcal{H}$ process independent

Elements of Proof

- Parameters L_0, M_0 related to memory:
  - L_0 large if forgetfulness slow
  - M_0 large if mixing slow
- Step 1:
  - Replace slow stage with a modification
  - Replace process with a block-independent process
  - Establish monopolarization
- Step 2:
  - Choose suitable L_0, M_0
  - Show negligible difference between step 1 replacements and actual process, slow stage
  - Implies main result