General Strong Polarization

Madhu Sudan
Harvard University

Joint work with Jaroslaw Blasiok (Harvard), Venkatesan Guruswami (CMU), Preetum Nakkiran (Harvard) and Atri Rudra (Buffalo)

This Talk: Martingales and Polarization

- [0,1]-Martingale: Sequence of r.v.s $X_0, X_1, X_2, \ldots$
  $$X_t \in [0,1], \qquad \mathbb{E}[X_{t+1} | X_0, \ldots, X_t] = X_t$$
- Well-studied in algorithms:
  - [MotwaniRaghavan, §4.4], [MitzenmacherUpfal, §12]
  - E.g., $X_t =$ Expected approx. factor of algorithm given first t random choices.
  - Usually study "concentration"
    - E.g. … whp $X_{\text{final}} \in [X_0 \pm \sigma] \Leftrightarrow \lim_{t\to\infty} X_t = N(X_0, \sigma^2)$
- Today: Polarization:
  - … whp $X_{\text{final}} \notin (\epsilon, 1-\epsilon) \Leftrightarrow \lim_{t\to\infty} X_t = \text{Bern}(X_0)$
- Why? – Polar codes …

Some examples

- $X_{t+1} = \begin{cases} X_t + 4^{-t} \text{ w. p. } \frac{1}{2} \\ X_t - 4^{-t} \text{ w. p. } \frac{1}{2} \end{cases}$ — Converges!
- $X_{t+1} = \begin{cases} X_t + 2^{-t} \text{ w. p. } \frac{1}{2} \\ X_t - 2^{-t} \text{ w. p. } \frac{1}{2} \end{cases}$ — Uniform on [0,1]
- $X_{t+1} = \begin{cases} \frac{3}{2} X_t \text{ w. p. } \frac{1}{2} & \text{if } X_t \leq \frac{1}{2} \\ \frac{1}{2} X_t \text{ w. p. } \frac{1}{2} & \text{if } X_t \leq \frac{1}{2} \end{cases}$ — Polarizes (weakly)!
- $X_{t+1} = \begin{cases} X_t^2 \text{ w. p. } \frac{1}{2} & \text{if } X_t \leq \frac{1}{2} \\ 2X_t - X_t^2 \text{ w. p. } \frac{1}{2} & \text{if } X_t \leq \frac{1}{2} \end{cases}$ — Polarizes (strongly)!

Main Result: Definition and Theorem

- Strong Polarization: (informally)

$$\Pr[X_t \in (\tau, 1-\tau)] \le \epsilon \text{ if } t \ge \max\{O(\log 1/\epsilon), o(\log 1/\tau)\}$$

formally $\forall \gamma > 0\ \exists \beta < 1, c\ s.t. \forall t\ \Pr[X_t \in (\gamma^t, 1-\gamma^t)] \le c \cdot \beta^t$

- Local Polarization:
  - Variance in the middle: $X_t \in (\tau, 1-\tau)$

$$\forall \tau > 0\ \exists \sigma > 0\ s.t. \forall t, X_t \in (\tau, 1-\tau) \Rightarrow Var[X_{t+1} | X_t] \ge \sigma$$

  - Suction at the ends: $X_t \notin (\tau, 1-\tau)$

$$\exists \theta > 0, \forall c < \infty, \exists \tau > 0\ s.t. X_t < \tau \Rightarrow \Pr\left[X_{t+1} < \frac{X_t}{c}\right] \ge \theta$$

- Theorem: Local Polarization ⇒ Strong Polarization.

Both definitions qualitative!

"low end" condition. Similar condition for high end

Proof (Idea):

- Step 1: The potential $\Phi_t \stackrel{\text{def}}{=} \min\{\sqrt{X_t}, \sqrt{1-X_t}\}$ decreases by constant factor in expectation in each step.
  - $\Rightarrow \mathbb{E}[\Phi_T] = \exp(-T)$
  - $\Rightarrow \Pr[X_T \geq \exp(-T/2)] \leq \exp(-T/2)$
- Step 2: Next T time steps, X_t plummets whp
  - 2.1: Say, If $X_t \leq \tau$ then $\Pr\left[X_{t+1} \leq \frac{X_t}{100}\right] \geq 1/2$.
  - 2.2: $\Pr[\exists t \in [T, 2T]\ s.t.\ X_t > \tau\ ] \leq X_T/\tau$ [Doob]
  - 2.3: If above doesn't happen $X_{2T} < 5^{-T}$ whp

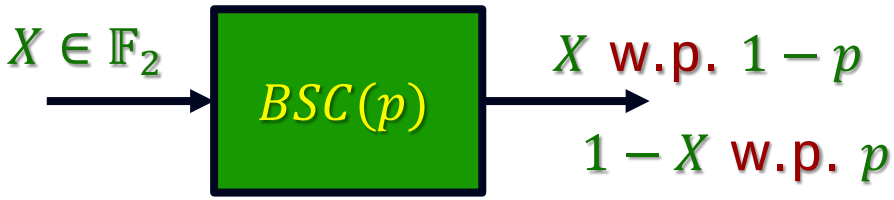
QED

Rest of the talk

- Background/Motivation:
  - The Shannon Capacity Challenge
  - Resolution by [Arikan'08], [Guruswami-Xia'13], [Hassani,Alishahi,Urbanke'13]: Polar Codes!
  - Key ideas:
    - Polar Codes & Arikan-Martingale.
    - Implications of Weak/Strong Polarization.
- Our Contributions: Simplification & Generalization
  - Local Polarization $\Rightarrow$ Strong Polarization
  - Arikan-Martingale Locally Polarizes (generally)

Shannon and Channel Capacity

- Shannon ['48]: For every (noisy) channel of communication, ∃ capacity C s.t. communication at rate $R < C$ is possible and $R > C$ is not.
  - Needs lot of formalization – omitted.
- Example:



Capacity?

∃ Encode: $\mathbb{F}_2^{k_n} \to \mathbb{F}_2^n$

$\lim_{n\to\infty} \frac{k_n}{n} \geq R$ (Rate)

  - Acts independently on bits
  - Capacity $= 1 - H(p)$; $H(p)$ = binary entropy!
  - $H(p) = p \cdot \log\frac{1}{p} + (1-p) \cdot \log\frac{1}{1-p}$

"Achieving" Channel Capacity

- Commonly used phrase. Many mathematical interpretations.
- Some possibilities:
  - How small can n be? Shannon '48 : $n = O(C - R)^{-2}$
  - Get $R > C - \epsilon$ with polytime algorithms?

    Forney '66 :time $= poly\left(n, 2^{\frac{1}{\epsilon^2}}\right)$
  - [Luby et al.'95] Make running time $\text{poly}\left(\frac{n}{\epsilon}\right)$?
    - Open till 2008
    - Arikan'08: Invented "Polar Codes" …
    - Final resolution: Guruswami+Xia'13, Hassani+Alishahi+Urbanke'13 – Strong analysis

Polar Codes and Polarization

- Arikan:
  - Defined Polar Codes, one for every t
  - Associated Martingale $X_0, \ldots, X_t, \ldots$
  - tth code: Length $n = 2^t$
  - If $\Pr[X_t \in (\gamma, 1-\delta)] \leq \epsilon$ then tth code is $(\epsilon+\delta)$-close to capacity, and
    - $\Pr[\text{Decode}(\text{Encode}(m) + error) \neq m] \leq n \cdot \gamma$
    - Need $\gamma = o\left(\frac{1}{n}\right)$ (strong polarization $\Rightarrow \gamma = neg(n)$)
    - Need $\epsilon = 1/n^{\Omega(1)}$ ($\Leftarrow$ strong polarization)

Context for today's talk

- [Arikan'08]: Martingale associated to general class of codes.
  - Martingale polarizes (weakly) implies code achieves capacity (weakly)
  - Martingale polarizes weakly for entire class.
- [GX'13, HAU'13]: Strong polarization for one specific code.
- Why so specific?
  - Bottleneck: Strong Polarization
- Rest of talk: Polar codes, Martingale and Strong Polarization

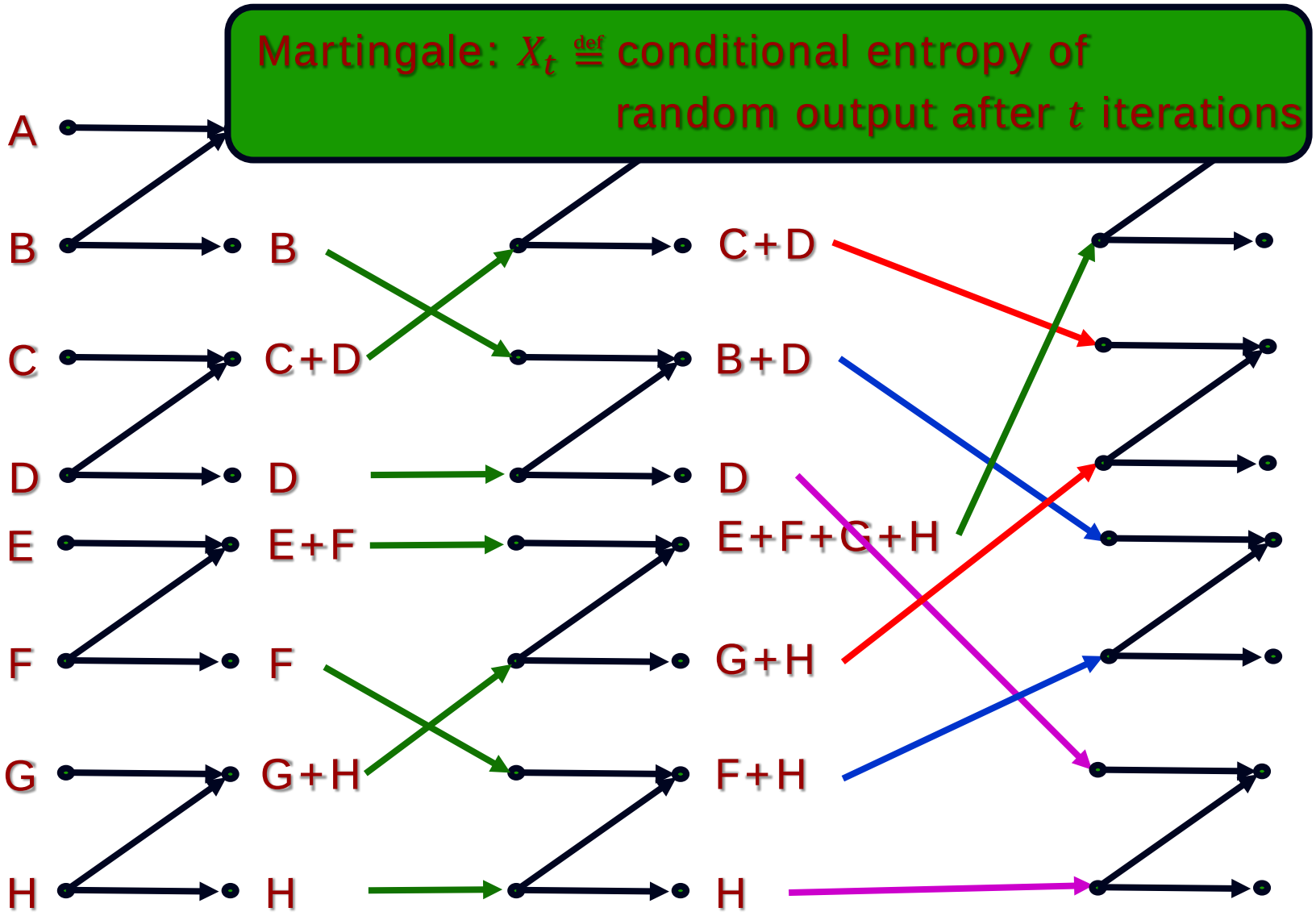
Lesson 0: Compression ⇒ Coding

- Defn: Linear Compression Scheme:
  - (H, D) for compression scheme for $\mathrm{bern}(p)^n$ if
    - Linear map $H: \mathbb{F}_2^n \to \mathbb{F}_2^m$
    - $\Pr_{Z \sim \mathrm{bern}(p)^n}[D(H \cdot Z) \neq Z] = o(1)$
    - Hopefully $\frac{m}{n} \leq H(p) + \epsilon$
- Compression ⇒ Coding
  - Let $H^\perp$ be such that $H \cdot H^\perp = 0$;
  - Encoder: $X \mapsto H^\perp \cdot X$
  - Error-Corrector: $Y = H^\perp \cdot X + Z \mapsto Y - D(H \cdot Y)$
    $= Y - D(H \cdot H^\perp \cdot X + H.Z) =_{w.p.\ 1-o(1)} H^\perp \cdot X$

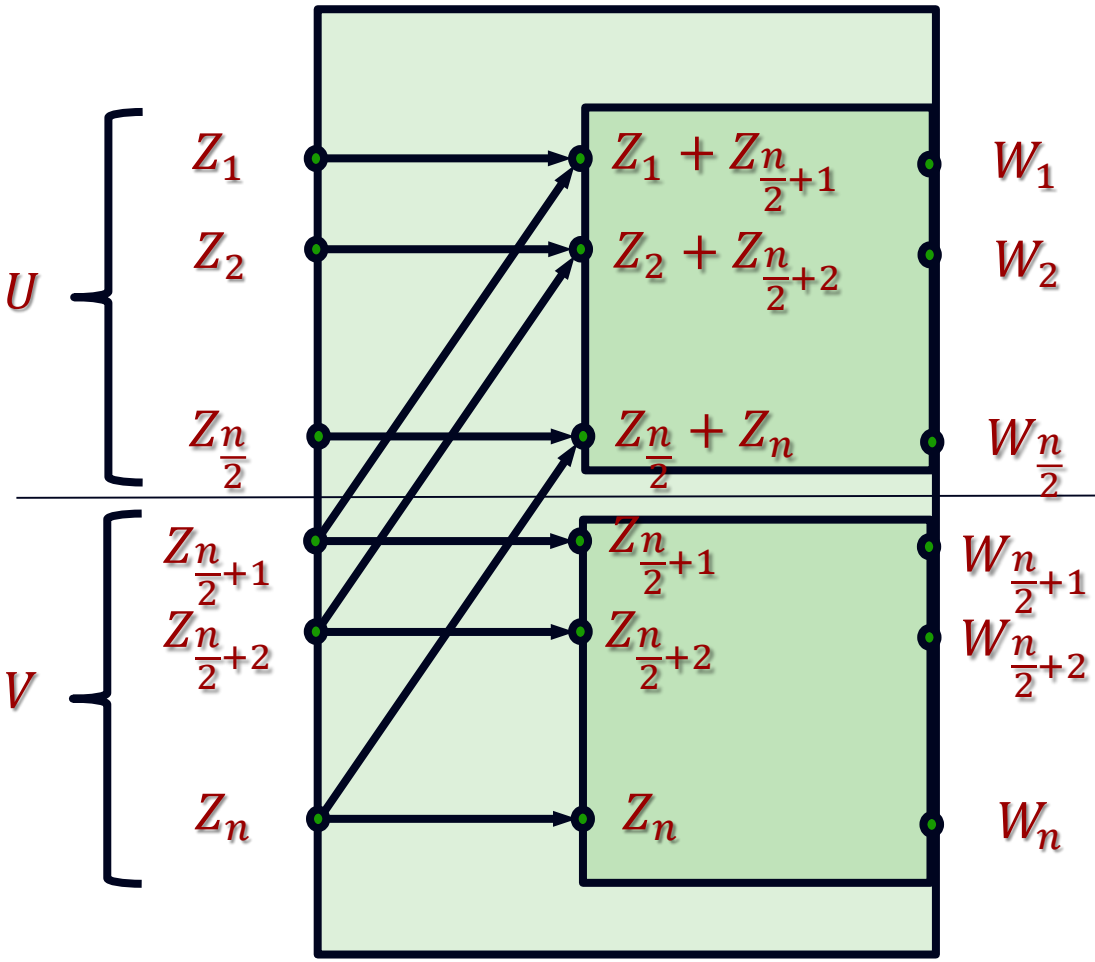
Question: How to compress?

- Arikan's key idea:
  - Start with 2×2 "Polarization Transform":
    $$(U, V) \to (U + V, V)$$
    - Invertible – so does nothing?
    - If U, V independent,
      - then $U + V$ "more random" than either
      - $V \mid U + V$ "less random" than either
  - Iterate (ignoring conditioning)
    - End with bits that are almost random, or almost determined (by others).
    - Output "totally random part" to get compression!

Iteration elaborated:



The Polarization Butterfly



$$P_n(U,V) = \left(P_{\frac{n}{2}}(U+V), P_{\frac{n}{2}}(V)\right)$$

Polarization $\Rightarrow \exists S \subseteq [n]$

$$H\left(W_{[n]-S} \middle| W_S\right) \to 0$$

$$|S| \leq (H(p) + \epsilon) \cdot n$$

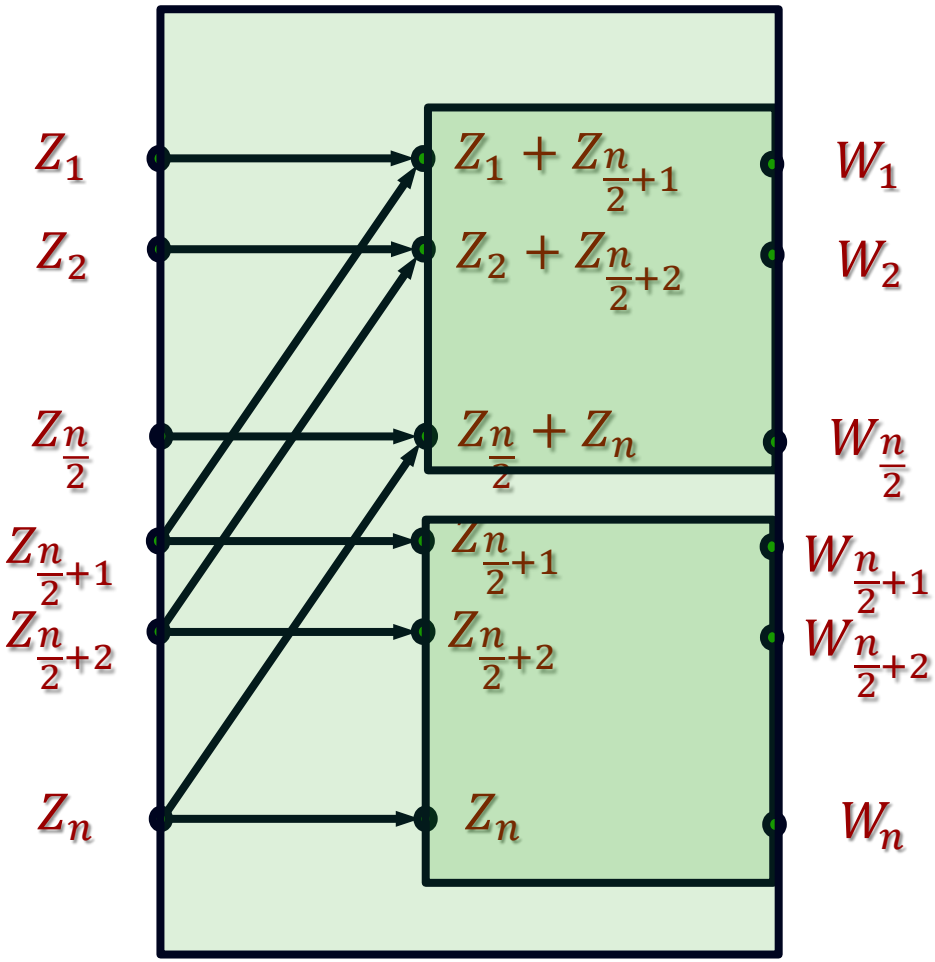
Compression $E(Z) = P_n(Z)_S$

Encoding time $= O(n \log n)$ (given S)

To be shown:

1. Decoding = $O(n \log n)$
2. Polarization!

The Polarization Butterfly: Decoding



Decoding idea:

Given $S, W_S = P_n(U,V)$

Compute $U+V \sim \text{Bern}(2p-2p^2)$

Determine $V \sim ?$

$$V|U+V \sim \prod_i \text{Bern}(r_i)$$

Key idea: Stronger induction!

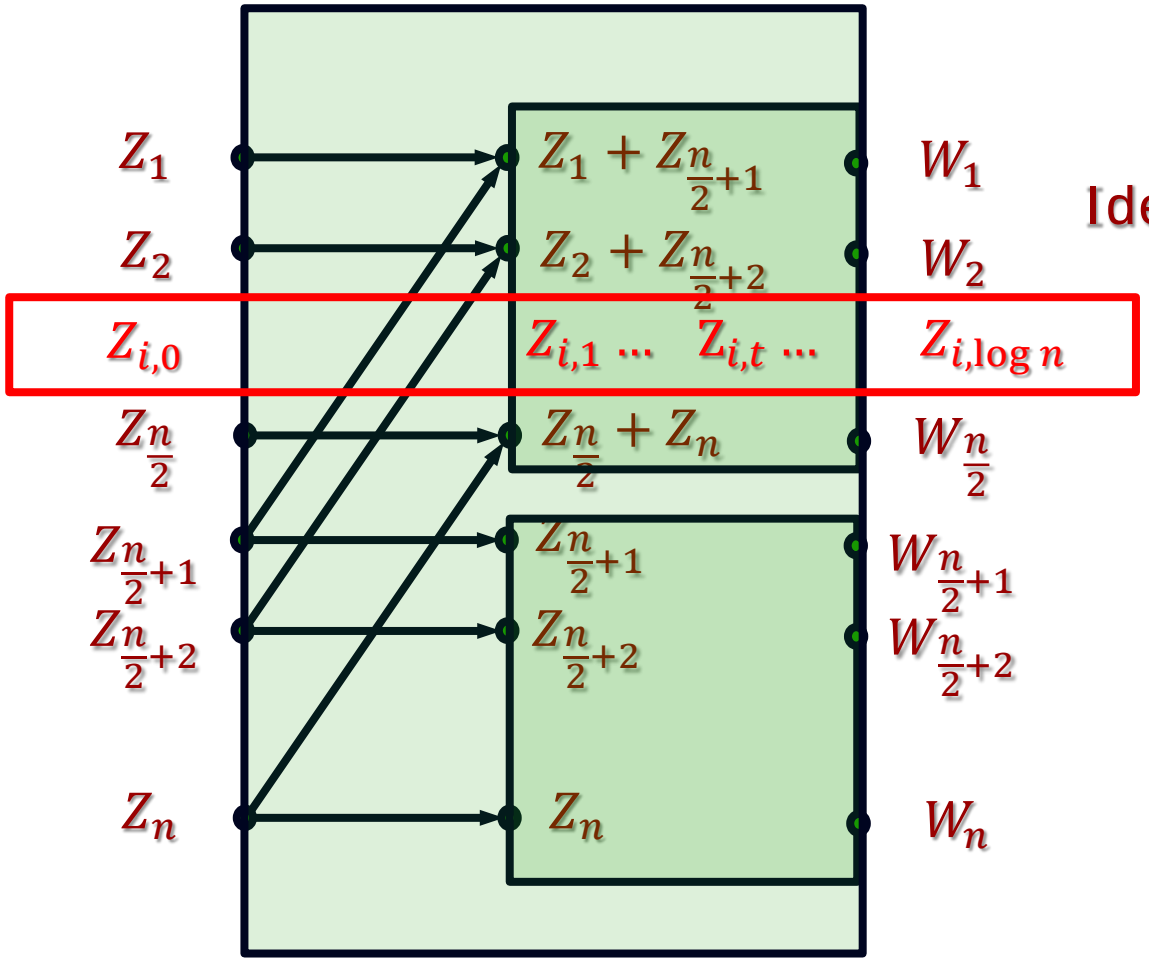
- non-identical product distribution

Decoding Algorithm:

Given $S, W_S = P_n(U,V), p_1, \dots, p_n$

- Compute $q_1, .. q_{\frac{n}{2}} \leftarrow p_1 \dots p_n$
- Determine $U+V = D\left(W_S^+, q_1 \dots q_{\frac{n}{2}}\right)$
- Compute $r_1 \dots r_{\frac{n}{2}} \leftarrow p_1 \dots p_n; U+V$
- Determine $V = D\left(W_S^-, r_1 \dots r_{\frac{n}{2}}\right)$

Polarization Martingale



Idea: Follow $X_t = H(Z_{i,t}|Z_{<i,t})$ for random i

X_t Martingale!

Polarizes?

Strongly?

⇐ Locally?

Local Polarization of Arikan Martingale

- Variance in the Middle:
  - Roughly: $\big(H(p), H(p)\big) \to (H(2p-2p^2), 2H(p) - H(2p-2p^2))$
    - $p \in (\tau, 1-\tau) \Rightarrow 2p - 2p^2$ far from p
      + continuity of $H(\cdot) \Rightarrow H(2p-2p^2)$ far from $H(p)$
- Suction:
  - High end: $H\left(\frac{1}{2} - \gamma\right) \to H\left(\frac{1}{2} - \gamma^2\right)$

    $$H\left(\tfrac{1}{2} - \gamma\right) = 1 - \Theta(\gamma^2) \Rightarrow 1 - \gamma^2 \to 1 - \Theta(\gamma^4)$$
  - Low end: $H(p) \approx p \log\frac{1}{p}; 2H(p) \approx 2p \log\frac{1}{p}$;

    $$H(2p-2p^2) \approx H(2p) \approx 2p \log\frac{1}{2p} \approx 2H(p) - 2p \approx \left(2 - \frac{1}{\log\frac{1}{p}}\right) H(p)$$
- Dealing with conditioning – more work (lots of Markov)

"New contributions"

- So far: Reproduced old work (simpler proofs?)
- Main new technical contributions:
  - Strong polarization of general transforms
    - E.g. $(U,V,W) \to (U+V,V,W)$?
  - Exponentially strong polarization [BGS'18]
    - Suction at low end is very strong!
    - Random $k \times k$ matrix yields $X_{t+1} \approx X_t^{k^{.99}}$ whp
  - Strong polarization of Markovian sources [GNS'19?]
    - Separation of compression of known sources from unknown ones.

Conclusions

- Polarization:
  - Important phenomenon associated with bounded random variables.
  - Not always bad ☹ ☺
  - Needs better understanding!
    - Recently also used in CS:
      - [Lovett Meka] – implicit
      - [Chattopadhyay-Hatami-Hosseini-Lovett] - explicit

Thank You!