General Strong Polarization

Madhu Sudan
Harvard University

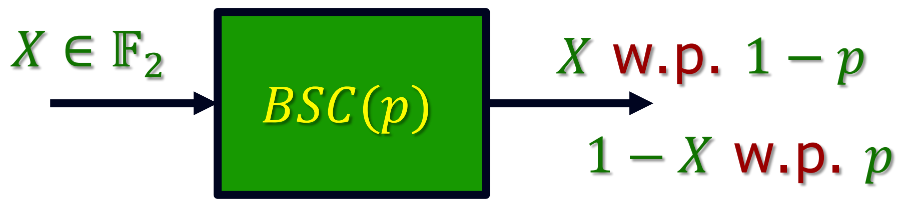
Based on joint works with Jaroslaw Blasiok (Harvard), Venkatesan Guruswami (CMU), Preetum Nakkiran (Harvard) and Atri Rudra (Buffalo)

Addendum: (After Alex's talk yesterday)

- Another talk on Polar Codes.
- Emphasis
  - BSC – "errors"
  - Focus on asymptotics and theorems!
  - … and proofs
  - … hopefully some teachable material

Shannon and Channel Capacity

- BSC(p):



  - Acts independently on bits
  - Capacity $= 1 - h(p)$; $h(p)$ = binary entropy!
  - $h(p) = p \cdot \log \frac{1}{p} + (1-p) \cdot \log \frac{1}{1-p}$
  - This talk: Price of communication at rate $R = C - \epsilon$
    - Smallest n, smallest running times.

"Achieving" Shannon Capacity

- How small can n be? Shannon '48: $n = \Theta\left(\frac{1}{\epsilon^2}\right)$; $\epsilon \stackrel{\text{def}}{=} C - R$
- Get $R > C - \epsilon$ with polytime algorithms?

  Forney '66 :time $= poly\left(n, 2^{\frac{1}{\epsilon^2}}\right)$
- Problem articulated by [Luby et al.'95]

  running time $\mathrm{poly}\left(\frac{n}{\epsilon}\right)$?

  (equiv. want block length $n = \mathrm{poly}\left(\frac{1}{\epsilon}\right)$?)
  - Open till 2008
  - Arikan'08: Invented "Polar Codes" …
  - Resolution of open question: Guruswami+Xia'13, Hassani+Alishahi+Urbanke'13 – Strong analysis

Polar Codes and Martingales

- Arikan: Defined Polar Codes, one for every integer t
- Associated "martingale" $X_0, \dots, X_t, \dots \quad X_t \in [0,1]$
- tth

> $X_0, X_1, \dots, X_t, \dots$ form a martingale if
> $$\forall t, \qquad \mathbb{E}[X_t | X_0, \dots, X_{t-1}] = X_{t-1}$$

  - tth code is $(\epsilon_t + \delta_t)$-close to capacity, and
  - $\Pr\left[\mathrm{Decode}\left(\mathrm{BSC}\left(\mathrm{Encode}(m)\right)\right) \neq m\right] \leq n \cdot \tau_t$
  - Need $\tau_t = o\left(\frac{1}{n}\right)$ or $\tau_t = \frac{1}{n^{\omega(1)}}$
  - Need $\epsilon_t, \delta_t = 1/n^{\Omega(1)}$

  } "Strong" Polarization

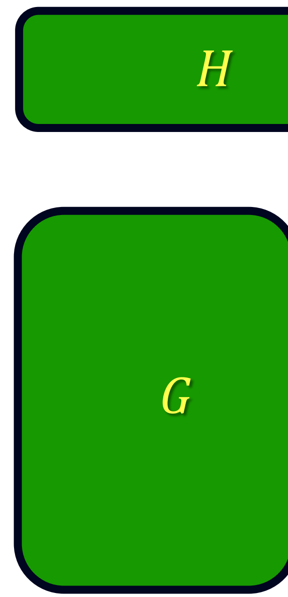
  - Arikan et al. $\tau = neg(n); \epsilon = o(1)$; [GX13,HAU13] ↑

Part II: Polar Codes
Encoding, Decoding, Martingale, Polarization

Lesson 0: Compression ⇒ Coding

- Defn: Linear Compression Scheme:
  - (M, D) form compression scheme for $\text{Bern}(p)^n$ if
    - Linear map $H: \mathbb{F}_2^n \to \mathbb{F}_2^m$
    - $\Pr_{Z \sim \text{Bern}(p)^n}[D(H \cdot Z) \neq Z] = o(1)$
    - Want: $\frac{m}{n} \leq h(p) + \epsilon$, D efficient
- Compression ⇒ Coding
  - Let G be such that $H \cdot G = 0$;
  - Encoder: $X \mapsto G \cdot X$
  - Error-Corrector: $Y = G \cdot X + Z \mapsto Y - D(H \cdot Y)$

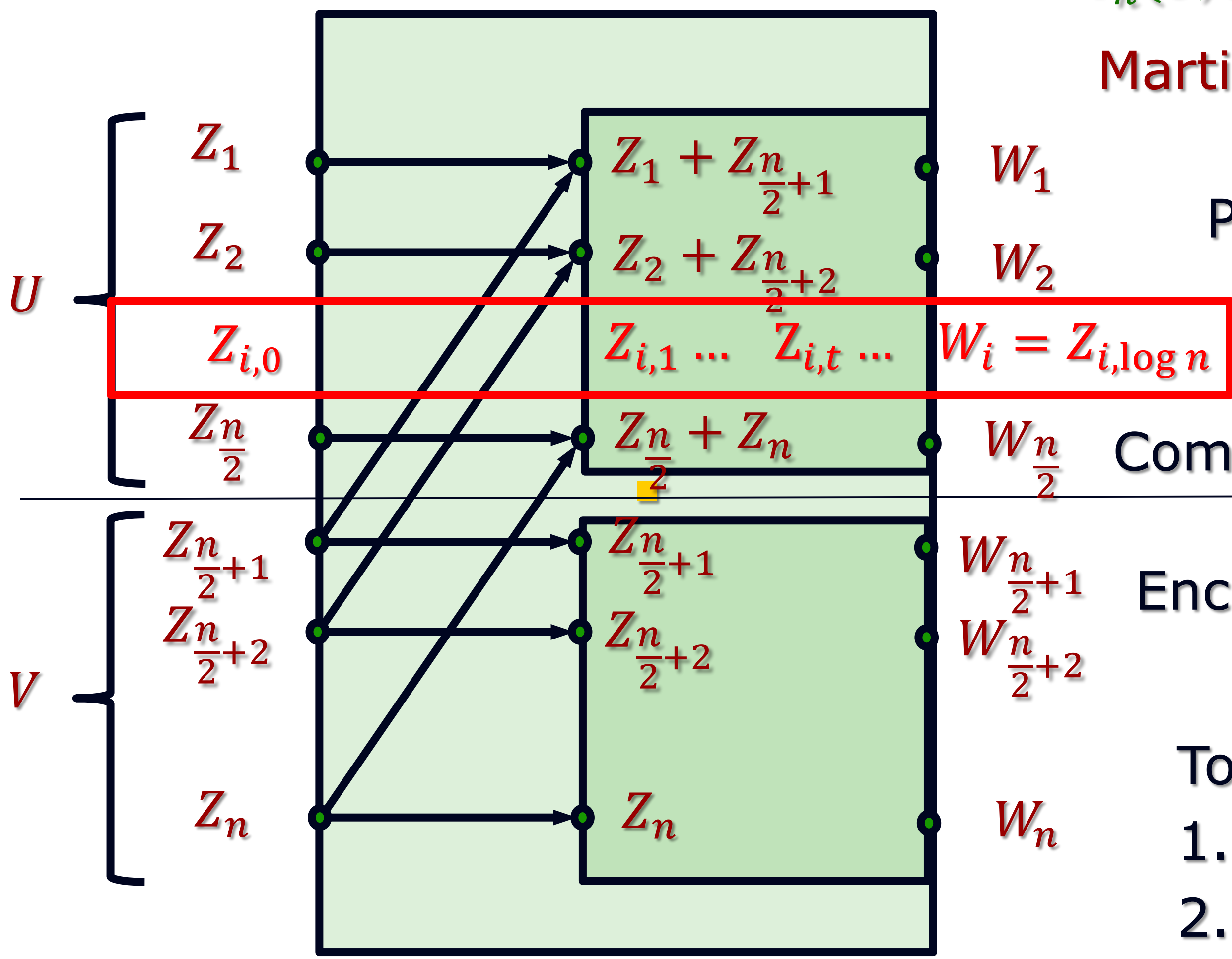
    $= Y - D(H \cdot G \cdot X + H \cdot Z) =_{w.p.\ 1-o(1)} G \cdot X$



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 "random part" to get compression!

The Polarization Butterfly



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

Martingale: $X_t = H\left(Z_{i,t} \middle| Z_{<i,t}\right)$

for random i

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

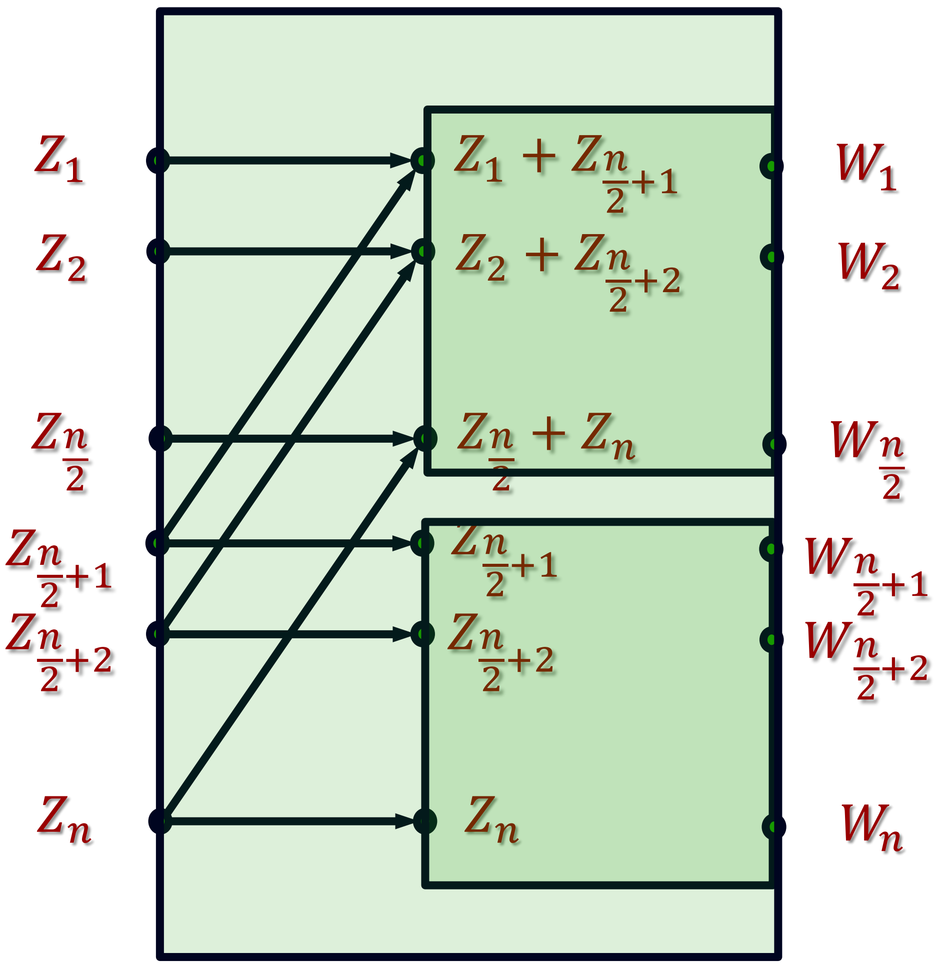
Encoding time $= O(n \log n)$

(given S)

To be shown:

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

The Polarization Butterfly: Decoding



Decoding idea:

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

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

Determine $V \sim ?$

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

Key idea: Stronger induction!

- non-identical product distribution

Decoding Algorithm:

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

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

Part III: Martingales, Polarization, Strong & Local

Martingales: Toy examples

- $X_{t+1} = \begin{cases} X_t + 2^{-t^2} \text{ w.p. } \frac{1}{2} \\ X_t - 2^{-t^2} \text{ w.p. } \frac{1}{2} \end{cases}$

  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]

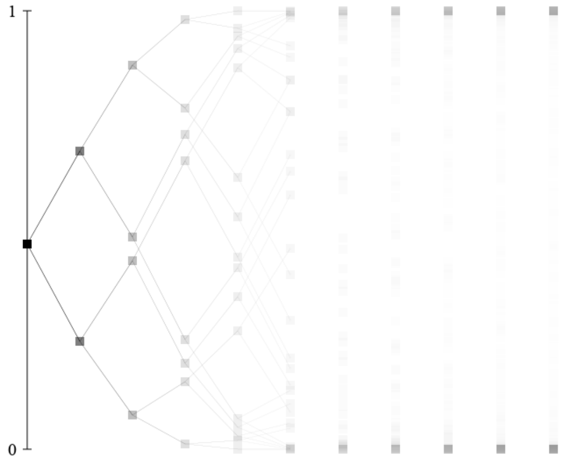
- $X_{t+1} = \begin{cases} \frac{3}{2} X_t \text{ w.p. } \frac{1}{2} & \text{if } X_t \leq \frac{1}{2} \\ \frac{1}{2} X_t \text{ w.p. } \frac{1}{2} & \text{if } X_t \leq \frac{1}{2} \end{cases}$

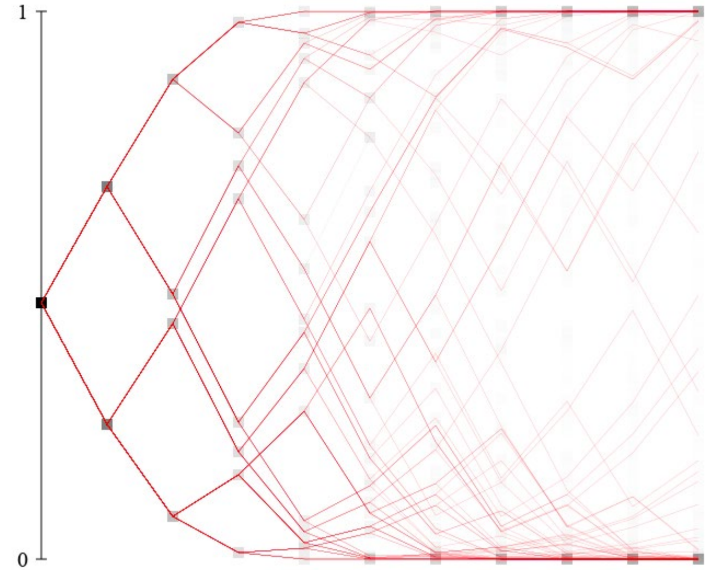
  Polarizes (weakly)!

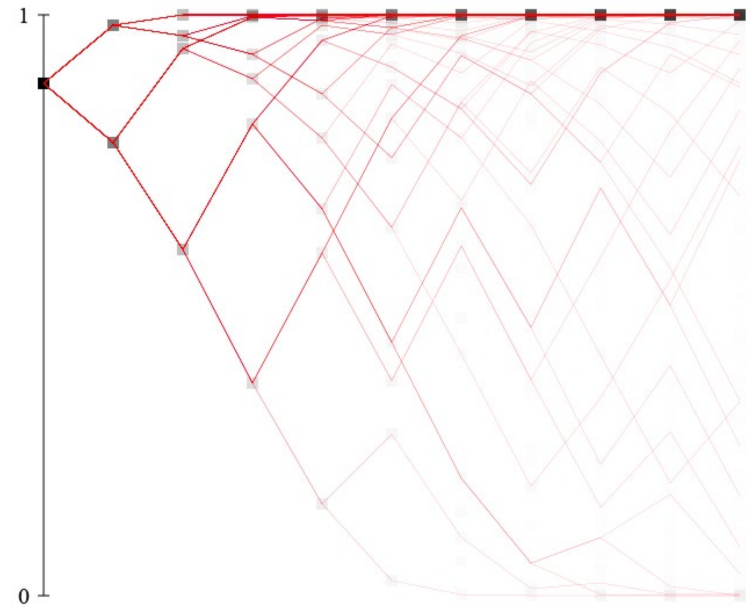
- $X_{t+1} = \begin{cases} X_t^2 \text{ w.p. } \frac{1}{2} & \text{if } X_t \leq \frac{1}{2} \\ 2X_t - X_t^2 \text{ w.p. } \frac{1}{2} & \text{if } X_t \leq \frac{1}{2} \end{cases}$

  Polarizes (strongly)!

Arikan Martingale







Issues:
Local behavior – well understood
Challenge: Limiting behavior

Main Result: Definition and Theorem

- Strong Polarization: (informally)

  $\Pr[X_t \in (\tau, 1-\tau)] \leq \epsilon$ if $\tau = 2^{-\omega(t)}$ and $\epsilon = 2^{-O(t)}$

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

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

    $\forall \tau > 0 \; \exists \sigma > 0 \; s.t. \forall t, X_t \in (\tau, 1-\tau) \Rightarrow Var[X_{t+1} | X_t] \geq \sigma$
  - Suction at the ends: $X_t \notin (\tau, 1-\tau)$

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

- 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

Local Polarization of Arikan Martingale

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

    $H\left(\frac{1}{2} - \gamma\right) = 1 - \Theta(\gamma^2) \Rightarrow 1 - \gamma^2 \to 1 - \Theta(\gamma^4)$
  - 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

- Importance of Strong Polarization!
- Generality of Strong Polarization!

- Some technical questions:
  - Best $\text{poly}\left(\frac{1}{\epsilon}\right)$?
  - Now turn to worst case errors? (with list-decoding)

Thank You!