Polar Codes: A tutorial

Madhu Sudan
Harvard University

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

- Not a survey … but a technical deep dive
- 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 = $1 - h(p)$; $h(p) = \text{binary entropy}$!
  - $h(p) = p \cdot \log \frac{1}{p} + (1 - p) \cdot \log \frac{1}{1-p}$
  - This talk: Price of communication at rate $R = C - \epsilon$
    - Smallest $n$, smallest running times, subject to $o_\epsilon(1)$ error

$X \in F_2$  
$X \text{ w.p. } 1 - p$  
$1 - X \text{ w.p. } p$
“Achieving” Shannon Capacity

- How small can $n$ be?  
  Shannon ‘48: $n = \Theta\left(\frac{1}{\epsilon^2}\right)$

- Get $R > C - \epsilon$ with polytime algorithms?
  Forney ‘66: $\text{time} = \text{poly}(n, 2^{\frac{1}{\epsilon^2}})$
“Achieving” Shannon Capacity

- How small can $n$ be?  
  Shannon ‘48: $n = \Theta \left( \frac{1}{\epsilon^2} \right)$

- Get $R > C - \epsilon$ with polytime algorithms?
  
    Forney ‘66: time = $\text{poly}(n, 2\epsilon^2)$

- Problem articulated by [Luby et al.’95]

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

  (equiv. want block length $n = \text{poly} \left( \frac{1}{\epsilon} \right)$?)
“Achieving” Shannon Capacity

- How small can $n$ be? Shannon ‘48: $n = \Theta\left(\frac{1}{\epsilon^2}\right)$
- Get $R > C - \epsilon$ with polytime algorithms? Forney ‘66: \(\text{time} = \text{poly}(n, 2^{1/\epsilon^2})\)
- Problem articulated by [Luby et al.’95]
  
  running time $\text{poly}\left(\frac{n}{\epsilon}\right)$?

  (equiv. want block length $n = \text{poly}\left(\frac{1}{\epsilon}\right)$?)

- Open till 2008
- Arikan’08: Invented “Polar Codes” ...
- Resolution of open question: Guruswami+Xia’13, Hassani+Alishahi+Urbanke’13 – Strong analysis
Outline of this talk

- BSC error correction via linear compression
- Polar Codes & Polarization
- Polarization & Martingales
- Polarization of the Arikan Martingale
- Algorithmics
Lesson 0: Compression ⇒ Coding

- Defn: Linear Compression Scheme:
  - \((M, D)\) form compression scheme for \(Bern(p)^n\) if
    - Linear map \(H: \mathbb{F}_2^n \to \mathbb{F}_2^m\)
    - \(\Pr_{Z \sim Bern(p)^n}[D(H \cdot Z) \neq Z] = o(1)\)
    - Want: \(\frac{m}{n} \leq h(p) + \epsilon, D\) efficient
Lesson 0: Compression ⇒ Coding

- Defn: Linear Compression Scheme:
  - \((M, D)\) form compression scheme for \(\text{Bern}(p)^n\) if
    - Linear map \(H : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^m\)
    - \(\Pr_{Z \sim \text{Bern}(p)^n} [D(H \cdot Z) \neq Z] = o(1)\)
    - 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: \(Y = G \cdot X + Z \mapsto Y - D(H \cdot Y) = Y - D(H \cdot G \cdot X + H \cdot Z) =_{\text{w.p.}} 1-o(1) G \cdot X\)
Outline of this talk

- BSC error correction via linear compression
- Polar Codes & Polarization
- Polarization & Martingales
- Polarization of the Arikan Martingale
- Algorithmics
Question: How to compress?

- Arikan’s key idea:
  - Start with $2 \times 2$ “Polarization Transform”:
    \[
    (U, V) \rightarrow (U + V, V)
    \]
Question: How to compress?

- Arikan’s key idea:
  - Start with $2 \times 2$ “Polarization Transform”:
    \[(U, V) \rightarrow (U + V, V)\]
  - Invertible – so does nothing?
  - If $U, V$ independent,
    - then $U + V$ “more random” than either
    - $V \mid U + V$ “less random” than either
Question: How to compress?

- Arikan’s key idea:
  - Start with $2 \times 2$ “Polarization Transform”: $\begin{pmatrix} U \\ V \end{pmatrix} \rightarrow \begin{pmatrix} U + V \\ V \end{pmatrix}$
  - 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:
Iteration elaborated:

A → A+B
B → B
C → C+D
D → D
Iteration elaborated:

\[ A \rightarrow A + B \rightarrow A + B + C + D \]

\[ B \rightarrow B \rightarrow C + D \rightarrow B + D \]

\[ C \rightarrow C + D \rightarrow B + D \]

\[ D \rightarrow D \rightarrow D \]
Iteration elaborated:
The Polarization Butterfly

\[ U \]
\[ Z_1 \]
\[ Z_2 \]
\[ Z_{n/2} \]
\[ Z_{n/2} + 1 \]
\[ Z_{n/2} + 2 \]
\[ Z_n \]

\[ V \]
\[ Z_1 + Z_{n/2+1} \]
\[ Z_2 + Z_{n/2+2} \]
\[ Z_{n/2} + Z_n \]
\[ Z_{n/2} + 1 \]
\[ Z_{n/2} + 2 \]
\[ Z_n \]
The Polarization Butterfly
The Polarization Butterfly

\[ Z_1 \]
\[ Z_2 \]
\[ \frac{Z_n}{2} \]
\[ \frac{Z_n}{2} + 1 \]
\[ \frac{Z_n}{2} + 2 \]

\[ W_1 \]
\[ W_2 \]
\[ W_{n/2} \]
\[ W_{n/2+1} \]
\[ W_{n/2+2} \]

\[ U \]
\[ V \]
The Polarization Butterfly

\[ P_n(U, V) = \left( \frac{1}{2} P_n(U + V), \frac{1}{2} P_n(V) \right) \]
The Polarization Butterfly

\[ P_n(U, V) = \left( \frac{P_n(U + V)}{2}, \frac{P_n(V)}{2} \right) \]

Polarization \( \Rightarrow \exists S \subseteq [n] \)

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

\[ |S| \leq (H(p) + \epsilon) \cdot n \]
The Polarization Butterfly

\[ P_n(U,V) = \left( \frac{P_n(U+V)}{2}, \frac{P_n(V)}{2} \right) \]

Polarization \( \Rightarrow \exists S \subseteq [n] \)

\[ H(W_{[n]-S} | W_S) \rightarrow 0 \]

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

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

To be shown:

1. Decoding = \( O(n \log n) \)
2. Polarization!
Outline of this talk

- BSC error correction via linear compression
- Polar Codes & Polarization
- Polarization & Martingales
- Polarization of the Arikan Martingale
- Algorithmics
Polar Codes and Martingales

- Defined Polar Codes, one for every integer $t$
- Associated “martingale” $X_0, ..., X_t, ...$ $X_t \in [0,1]$
- $t$th code: Length $n = 2^t$

$X_0, X_1, ..., X_t, ...$ form a martingale if

$$\forall t, \quad \mathbb{E}[X_t | X_0, ..., X_{t-1}] = X_{t-1}$$
The Polarization Butterfly

\[ P_n(U, V) = \left( \frac{P_n(U + V)}{2}, \frac{P_n(V)}{2} \right) \]

Polarization \( \Rightarrow \exists S \subseteq [n] \)
\[ H(W_{[n]-S} \mid W_S) \rightarrow 0 \]
\[ |S| \leq (H(p) + \varepsilon) \cdot n \]
Compression \( E(Z) = P_n(Z)_S \)

To be shown:
1. Decoding = \( O(n \log n) \)
2. Martingale Polarization
The Polarization Butterfly

\[ P_n(U, V) = \left( \frac{P_n(U + V)}{2}, \frac{P_n(V)}{2} \right) \]

Martingale: \[ X_t = H(Z_{i,t} | Z_{<i,t}) \]

for random \( i \)

Polarization \( \Rightarrow \) \( \exists S \subseteq [n] \)

\[ H(W_{[n]-S} | W_S) \rightarrow 0 \]

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

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

To be shown:

1. Decoding = \( O(n \log n) \)
2. Martingale Polarization
Polar Codes and Martingales

- Defined Polar Codes, one for every integer $t$
- Associated “martingale” $X_0, \ldots, X_t, \ldots$
- $t$th code: Length $n = 2^t$
Polar Codes and Martingales

- Defined Polar Codes, one for every integer $t$
- Associated “martingale” $X_0, \ldots, X_t, \ldots$
- $t$th code: Length $n = 2^t$
  - If $\Pr[X_t \in (\tau_t, 1 - \delta_t)] \leq \epsilon_t$ then
Polar Codes and Martingales

- Defined Polar Codes, one for every integer $t$
- Associated “martingale” $X_0, \ldots, X_t, \ldots$
- $t$th code: Length $n = 2^t$
  - If $\Pr[X_t \in (\tau_t, 1 - \delta_t)] \leq \epsilon_t$ then
  - $t$th code is $(\epsilon_t + \delta_t)$-close to capacity, and
  - $\Pr[\text{Decode}(\text{BSC}(\text{Encode}(m))) \neq m] \leq n \cdot \tau_t$
Polar Codes and Martingales

- Defined Polar Codes, one for every integer \( t \)
- Associated “martingale” \( X_0, \ldots, X_t, \ldots \)
- \( t \)th code: Length \( n = 2^t \)
  - If \( \Pr[X_t \in (\tau_t, 1 - \delta_t)] \leq \epsilon_t \) then
    - \( t \)th code is \((\epsilon_t + \delta_t)\)-close to capacity, and
  - \( \Pr\left[\text{Decode \( \text{BSC(Encode}(m)) \neq m \right] \leq n \cdot \tau_t \)
- Need \( \tau_t = o\left(\frac{1}{n}\right) \) or \( \tau_t = \frac{1}{n^{\omega(1)}} \)
- Need \( \epsilon_t, \delta_t = 1/n^{\Omega(1)} \)
- Arikan et al. \( \tau, \delta = \text{neg}(n); \epsilon = o(1); \)
- \([\text{GX13, HAU13}] \uparrow\)
Outline of this talk

- BSC error correction via linear compression
- Polar Codes & Polarization
- Polarization & Martingales
- Polarization of the Arikan Martingale
- Algorithmics
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 $[0,1]$

- $X_{t+1} = \begin{cases} \frac{3}{2} X_t \text{ w.p. } \frac{1}{2} \text{ if } X_t \leq \frac{1}{2} \\ \frac{1}{2} X_t \text{ w.p. } \frac{1}{2} \text{ if } X_t > \frac{1}{2} \end{cases}$
- Polarizes (weakly)!

- $X_{t+1} = \begin{cases} X_t^2 \text{ w.p. } \frac{1}{2} \text{ if } X_t \leq \frac{1}{2} \\ 2X_t - X_t^2 \text{ w.p. } \frac{1}{2} \text{ if } X_t > \frac{1}{2} \end{cases}$
- Polarizes (strongly)!
**Arikan Martingale**

**Issues:**
Local behavior – well understood
Challenge: Limiting behavior
Main Result: Definition and Theorem

- **Strong Polarization**: (informally)

- **Local Polarization**:

- **Theorem**: Local Polarization $\Rightarrow$ Strong Polarization.
Main Result: Definition and Theorem

- **Strong Polarization:** (informally)
  \[ \Pr[X_t \in (\tau, 1-\tau)] \leq \epsilon \text{ if } \tau = 2^{-\omega(t)} \text{ and } \epsilon = 2^{-O(t)} \]
  formally \[ \forall \gamma > 0 \exists \beta < 1, \text{ s.t. } \forall t \Pr[X_t \in (\gamma^t, 1-\gamma^t)] \leq c \cdot \beta^t \]

- **Local Polarization:**
Main Result: Definition and Theorem

- **Strong Polarization: (informally)**
  \[ \Pr[X_t \in (\tau, 1 - \tau)] \leq \epsilon \text{ if } \tau = 2^{-\omega(t)} \text{ and } \epsilon = 2^{-O(t)} \]
  formally \( \forall \gamma > 0 \exists \beta < 1, \text{ s.t. } \forall t \Pr[X_t \in (\gamma^t, 1 - \gamma^t)] \leq c \cdot \beta^t \)

- **Local Polarization:**
Main Result: Definition and Theorem

- **Strong Polarization:** (informally)
  \[ \Pr[X_t \in (\tau, 1-\tau)] \leq \epsilon \text{ if } \tau = 2^{-\omega(t)} \text{ and } \epsilon = 2^{-o(t)} \]
  formally \( \forall \gamma > 0 \exists \beta < 1, \text{s.t. } \forall t \Pr[X_t \in (\gamma^t, 1-\gamma^t)] \leq c \cdot \beta^t \)

- **Local Polarization:**
  - **Variance in the middle:** \( X_t \in (\tau, 1-\tau) \)
    \( \forall \tau > 0 \exists \sigma > 0 \text{ s.t. } \forall t, X_t \in (\tau, 1-\tau) \Rightarrow \text{Var}[X_{t+1} | X_t] \geq \sigma \)
  - **Suction at the ends:** \( X_t \notin (\tau, 1-\tau) \)
    \( \exists \theta > 0, \forall c < \infty, \exists \tau > 0 \text{ s.t. } X_t < \tau \Rightarrow \Pr\left[ X_{t+1} < \frac{X_t}{c} \right] \geq \theta \)

“low end” condition. Similar condition for high end.
Main Result: Definition and Theorem

- **Strong Polarization:** (informally)
  \[ \Pr[X_t \in (\tau, 1 - \tau)] \leq \epsilon \text{ if } \tau = 2^{-\omega(t)} \text{ and } \epsilon = 2^{-O(t)} \]
  formally \( \forall \gamma > 0 \exists \beta < 1, c \text{ s.t. } \forall t \Pr[X_t \in (\gamma^t, 1 - \gamma^t)] \leq c \cdot \beta^t \)

- **Local Polarization:**
  - **Variance in the middle:** \( X_t \in (\tau, 1 - \tau) \)
    \( \forall \tau > 0 \exists \sigma > 0 \text{ s.t. } \forall t, X_t \in (\tau, 1 - \tau) \Rightarrow \text{Var}[X_{t+1} | X_t] \geq \sigma \)
  - **Suction at the ends:** \( X_t \notin (\tau, 1 - \tau) \)
    \( \exists \theta > 0, \forall c < \infty, \exists \tau > 0 \text{ s.t. } X_t < \tau \Rightarrow \Pr\left[ X_{t+1} < \frac{X_t}{c} \right] \geq \theta \)

Both definitions qualitative!
Main Result: Definition and Theorem

- **Strong Polarization:** (informally)
  \[
  \Pr[X_t \in (\tau, 1 - \tau)] \leq \varepsilon \text{ if } \tau = 2^{-\omega(t)} \text{ and } \varepsilon = 2^{-O(t)}
  \]
  formally \(\forall \gamma > 0 \exists \beta < 1, c \text{ s.t. } \forall t \Pr[X_t \in (\gamma^t, 1 - \gamma^t)] \leq c \cdot \beta^t\)

- **Local Polarization:**
  - Variance in the middle: \(X_t \in (\tau, 1 - \tau)\)
    \(\forall \tau > 0 \exists \sigma > 0 \text{ s.t. } \forall t, X_t \in (\tau, 1 - \tau) \Rightarrow \text{Var}[X_{t+1} | X_t] \geq \sigma\)
  - Suction at the ends: \(X_t \notin (\tau, 1 - \tau)\)
    \(\exists \theta > 0, \forall c < \infty, \exists \tau > 0 \text{ s.t. } X_t < \tau \Rightarrow \Pr \left[ X_{t+1} < \frac{X_t}{c} \right] \geq \theta \)

- **Theorem:** Local Polarization \(\Rightarrow\) Strong Polarization.
Proof (Idea):

- **Step 1**: The potential \( \Phi_t \overset{\text{def}}{=} \min\{\sqrt{X_t}, \sqrt{1 - X_t}\} \) 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**: \( \Pr[\exists t \in [T, 2T] \text{ s.t. } X_t > \tau] \leq X_T/\tau \) [Doob]
  
  - **2.3**: If above doesn’t happen \( X_{2T} < 5^{-T} \) whp

QED
Local Polarization of Arikan Martingale

- **Variance in the Middle:**
  - Roughly: \((H(p), H(p)) \rightarrow (H(2p - 2p^2), 2H(p) - H(2p - 2p^2))\)
  - \(p \in (\tau, 1 - \tau) \Rightarrow 2p - 2p^2 \text{ far from } p\)
  - + continuity of \(H(\cdot) \Rightarrow H(2p - 2p^2) \text{ far from } H(p)\)

- **Suction:**
  - **High end:** \(H \left( \frac{1}{2} - \gamma \right) \rightarrow H \left( \frac{1}{2} - \gamma^2 \right)\)
    
    \[H \left( \frac{1}{2} - \gamma \right) = 1 - \Theta(\gamma^2) \Rightarrow 1 - \gamma^2 \rightarrow 1 - \Theta(\gamma^4)\]
  - **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)**
Outline of this talk

- BSC error correction via linear compression
- Polar Codes & Polarization
- Polarization & Martingales
- Polarization of the Arikan Martingale
- Algorithmics
Algorithmics: Overview

- Encoding: Manipulate bits
  - \( P_n(U,V) = \left( \frac{P_n(U+V)}{2}, \frac{P_n(V)}{2} \right) \Rightarrow O(n \log n) \)

- Decoding: Manipulate distributions on bits
  - Will see shortly

- Constructing the code (the set \( S \)): Manipulates conditional distributions = distribution over distributions.
  - Will say a few words.
The Polarization Butterfly: Decoding

Decoding idea:
Given $s, W_s = P_n(U, V)$
Compute $U + V \sim \text{Bern}(2p - 2p^2)$
Determine $V \sim ?$
The Polarization Butterfly: Decoding

Decoding idea:
Given \( S, W_S = P_n(U, V) \)
Compute \( U + V \sim \text{Bern}(2p - 2p^2) \)
Determine \( V \sim ? \)

Decoding Algorithm:
Given \( S, W_S = P_n(U, V), p_1, ..., p_n \)
- Compute \( q_1, ..., q_n \leftarrow p_1, ..., p_n \)
- Determine \( U + V = D(W_S^+, q_1, ..., q_n) \)
- Compute \( r_1, ..., r_n \leftarrow p_1, ..., p_n; U + V \)
- Determine \( V = D(W_S^-, r_1, ..., r_n) \)

Key idea: Stronger induction!
- non-identical product distribution
Constructing the set $S$

- At a core level, need to know $H(Z_{i,t}|Z_{<i,t})$
- Contrast to decoding: Used $H(Z_{i,t}|z_{<i,t})$
Constructing the set $S$

- At a core level, need to know $H(Z_{i,t} | Z_{<i,t})$
  - Contrast to decoding: Used $H(Z_{i,t} | Z_{<i,t})$
  - Much more challenging:
    - Need to know $Z_i | Z_{<i}$ for exponentially many $z_{<i}$
Constructing the set $S$

- At a core level, need to know $H(Z_{i,t} | Z_{<i,t})$
  - Contrast to decoding: Used $H(Z_{i,t} | z_{<i,t})$
- Much more challenging:
  - Need to know $Z_i | z_{<i}$ for exponentially many $z_{<i}$
  - Key observations in compressing this:
    - For every $p \in [0,1]$ only need to know how many $z_{<i}$ lead $Z_i \sim \text{Bern}(p)$ (not which ones).
    - Can also discretize the $p$'s.
- Many further details ... but in conclusion $S$ can be computed in poly time, even when $\tau_t \approx 2^{-c^t}$
Conclusions

- Importance of Strong Polarization!
- Generality of Strong Polarization!
  - (Can even get exponentially small error; Strong polarization happens whenever martingale polarizes weakly ...; can apply to Markovian sources/errors)
- Some technical questions:
  - Best poly \(\left(\frac{1}{\epsilon}\right)\) ? Resolved recently by [Guruswami-Riazanov-Ye]: \(n = \delta^{-2.0001}\)
  - Now need to turn to worst case errors? (with list-decoding)
Thank You!