LECTURE 16
Today

**Polar Codes**
- Motivation: Shannon & Gap to Capacity
- Construction:
  0. Reduction to Linear Compression
  1. Polarizing Transformation + Inf. Th. basics
  2. Polarization + Theorem
  3. Encoding + Decoding
  4. Proof of Polarization Theorem.

(some skipped)

Admin
- PS4 Due this week
- Weekly reports every week
- PS5 out tomorrow
- "Practice PS6" might be released later
- 4 best PS out of 6.
- Ask if you have questions.

Use Chat during lectures!
**Motivation**

![Diagram]

\[ P_k \left( m^* = m \right) \geq 1 - o(1) \]

- Can do this with rate \( \left( \frac{k}{m} \right) \rightarrow 1 - H(p) \)
  - \( p \) = Bit flip Prob.
- If we want to be at rate \( 1 - H(p) - \varepsilon \) then can achieve error prob \( \exp(-\varepsilon' n) \)
- Shannon - Non constructive
- \( \text{PS3} \) ? - Can make this poly time ....
  - Concatenation "disappeared" the problem?
- Running are now poly \( (n) \)
  - But there is a constant in front which depends on \( \varepsilon \).
- e.g. \( 2^{1/\varepsilon^2} \cdot n^2 \) needed because blocks are of size \( 1/\varepsilon^2 \) ... exponential in block.
**Problem Formulation:** Given $G$, determine $R, n$

\[ s.t.: \quad \frac{R}{n} \geq 1 - H(p) - \varepsilon \]

\[ \exists E_k: \mathbb{F}_2^k \rightarrow \mathbb{F}_2^k \]

\[ D_k: \mathbb{F}_2^n \rightarrow \mathbb{F}_2^k \]

\[ s.t. \quad \Pr_{m, b_{sc}} \left[ D \left( b_{sc} \left( E(m) \right) \right) \neq m \right] \leq 1 - o(1) \]

& Running time of $E_k, D_k$ are $\text{poly} \left( \frac{1}{\varepsilon} \right) \times X$

**History:**
- Q. raised by Luby, Mitzenmacher, Shokrollahi, Spielman '95
- 2008: Arikan - Proposed Polar Codes
- 2013: - Guruswami, Xia
- Hassani, Alishahi, Urbanke \{ resolved using Polar codes. \}
Codes specified by $G = \begin{bmatrix} \end{bmatrix}_k$,

or by

$H = \begin{bmatrix} \end{bmatrix}_n$

$G$ is generator of a good code (correctors)

$\implies H$ is a good linear compressor of Bern($p$)?

Bern($p$) = $Z = 0$ w.p. $1-p$

$= 1$ w.p. $p$

Bern($p$)$^n = n$ independent copies $Z_1...Z_n$

$Z_i \sim$ Bern($p$).

Is $H$ a good compressor?
If $E$ compresses $B_e^n$ if

1. $H = nxm$ with $m \leq (H(p)+\varepsilon)n$

2. $\exists$ Decoding alg $D$

\[
P_Z\left[D(\hat{Z}H) = Z\right] = o(1)
\]

Compression $Z$

3. $D$ should run in time $\text{poly}(\frac{1}{\varepsilon})$.

\[
\begin{align*}
\text{Enc} & \quad \rightarrow \quad \text{Channel} \quad \rightarrow \quad XG + Z \quad \rightarrow \quad \text{Can get} \\
X & \quad \rightarrow \quad XG \quad \rightarrow \quad \underbrace{XG + Z}_{Y} \quad \rightarrow \quad \underbrace{Z}_{X \text{ whp}}
\end{align*}
\]

$Y \cdot H = XGH + ZH = ZH$

$D(ZH) = \hat{Z} = Z \text{ whp}$

$Y - \hat{Z} = \text{whp our codeword}$.

\[
\begin{align*}
\text{Aside: Linear compression equivalent to our problem if we want linear code.}
\end{align*}
\]

Rest of Polar Coding: Linear Compression.
- Two bits & compress them:
  \[(U, V) \rightarrow (U+V, V)\]

- Move entropy around
  \[U+V \text{ more "uncertain" than } U \text{ or } V\]
  \[H(U+V) > H(U)\]
- \(V\) no more/less entropic \(V\)?
  "Conditional Entropy" \[H(V \mid U+V) < H(V \mid U) = H(V)\]

**Entropy**: of random variable \(X\) dist. on \([M]\)
- w/ \(P[X=i] = P_i\)

\[H(X) = \sum_{i=1}^{M} P_i \log_2 \frac{1}{P_i}\]

Entropy says how effectively we can compress
\(n\) ind. copies of \(X\); amortized

\([\text{Decoding } X_1, \ldots, X_n, X_i \sim X \text{ i.i.d. takes roughly } H(X) \cdot n \text{ bits }\]
Suppose $U, V \sim \text{Bern}(0.01)$

- $U + V \sim \text{Bern}(0.01999)$

- $U + V = 0 \Rightarrow Pr[U=1 \mid U+V=0] = 0.000101 \ldots$
  $\Rightarrow Pr[\neg U] \approx 0.98$

- $U + V = 1 \Rightarrow Pr[U=1 \mid U+V=1] = \frac{1}{2}$
  $\Rightarrow Pr[\neg U] = 0.02$

But in expectation $U$ will be "less random"

Given $U+V$.

---

Conditional Entropy

- $(X, Y)$ jointly dist. over $[M] \times [N]$

  $Pr[X = i \land Y = j] = P_{ij}$; $P_x$ marginal on $X$

- $H(Y \mid X) = \mathbb{E}_{x \sim P_x} \left[ H(Y \mid X = x) \right]$

---

1. *Chain Rule*

$H(X, Y) = H(X) + \underline{\text{bits needed to describe \( Y \) if you know \( X \)}}$

2. *Conditioning "reduces" entropy: $H(Y \mid X) \leq H(Y)$.
- f: is a one-to-one function
  \[ H(x) = H(f(x)) \]

- \[ H(u,v) = H(u+v, v) \] - ①

- (By calc/Exercise): \[ H(u+v) > H(u) \] - ②

- \[ H(v|u+v) = H(u+v, v) - H(u+v) \] by ①

- Chain Rule

- \[ = H(u,v) - H(u+v) \]

- \[ = H(u,v) - H(u) \]

- \[ = H(v) \] [since \( V \) & \( U \) independent].

**Polar coding idea**

- Let's iterate this process many times, moving conditional entropies around.

- "Polarization": At end every bit will conditional Entropy close to 0, or 1.

- Compression: throw away all bits with 0 entropy.
\[ H(W_i | W_1, \ldots, W_{i-1}) = \text{very close to 0} \quad \text{or very close to 1} \]

\[ S = \sum_i H(W_i | W_{<i}) \rightarrow 1 \]

Compression of \( Z = W|_S \)

1. This is linear

2. \( |S| \approx H(p) \cdot n \rightarrow \text{follows from all} \quad \begin{cases} \text{entropies are 0 or 1} \end{cases} \)

3. Given \( W|_S \) can compute \( W \) and then \( Z \) efficiently.

All together \( \Rightarrow \text{Polarization proves good compressor} \Rightarrow \text{gives our theorem.} \)