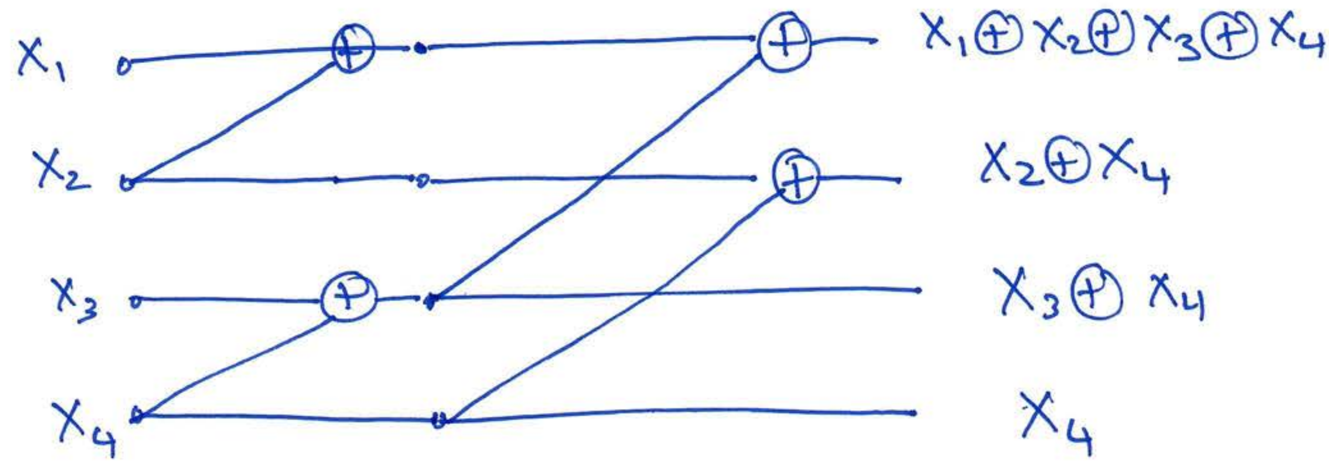
CS229r - LECTURE 16

3/23/2017

TODAY: POLAR CODES - II

- Error-correction (correcting my errors from last lecture ☺)
- Review Of Polar Codes
- *Decompression Algorithm
- Polarization Speed

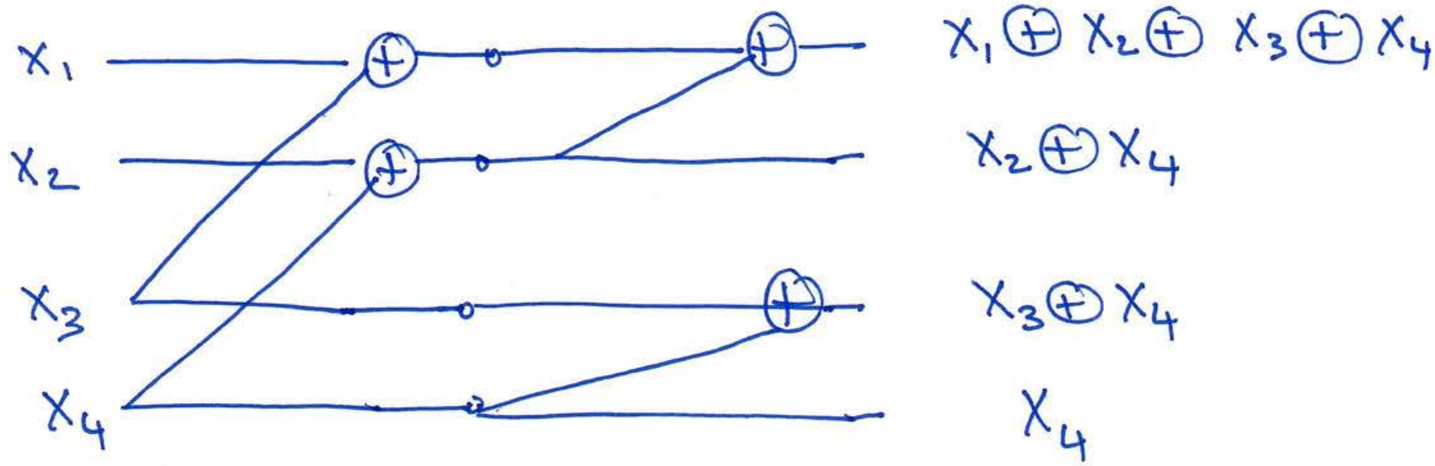
Last Lecture: ~~compress~~ Use following to "polarize"



This is WRONG

Conceptually, though not technically.

Correct Picture



But inputs → outputs is same; so what is different? Intermediate Nodes + Analysis!!

Recall eventual goal

Build linear circuit (such as above) $(X_1 .. X_n) \mapsto (Y_1 .. Y_n)$

s.t. for most i,

$$H(Y_i \mid Y_1 .. Y_{i-1}) \text{ is close to } 0 \text{ or } 1.$$

Analysis based on following idea:

At intermediate stage we may have computed A, B (linear forms) in $X_1 .. X_n$ & at next stage we produce

$$(A \oplus B,\ B).$$

But what polarizes are <u>conditional entropies</u>.

- So we have $H(A|C)$ & $H(B|D)$ are equa
for some variables C & D.

- But what conditional entropies in output should we measure?

- And how do we know that these correspond (at final layer) to

$$H(Y_i \mid \underbrace{Y_0 \cdots Y_{i-1}}_{\text{all previous outputs.}})$$

- Need to draw picture carefully.

will arrange it such that

(1) D is independent of (A, C)

(2) C is independent of (B, D)

(3) C & D are both "above" <u>A</u> & <u>B</u>.

(so at least all entropies we prove to be small are small when conditioned by <u>all</u> variables above)

(4) Remaining above variables independent of (A, C, B, D).

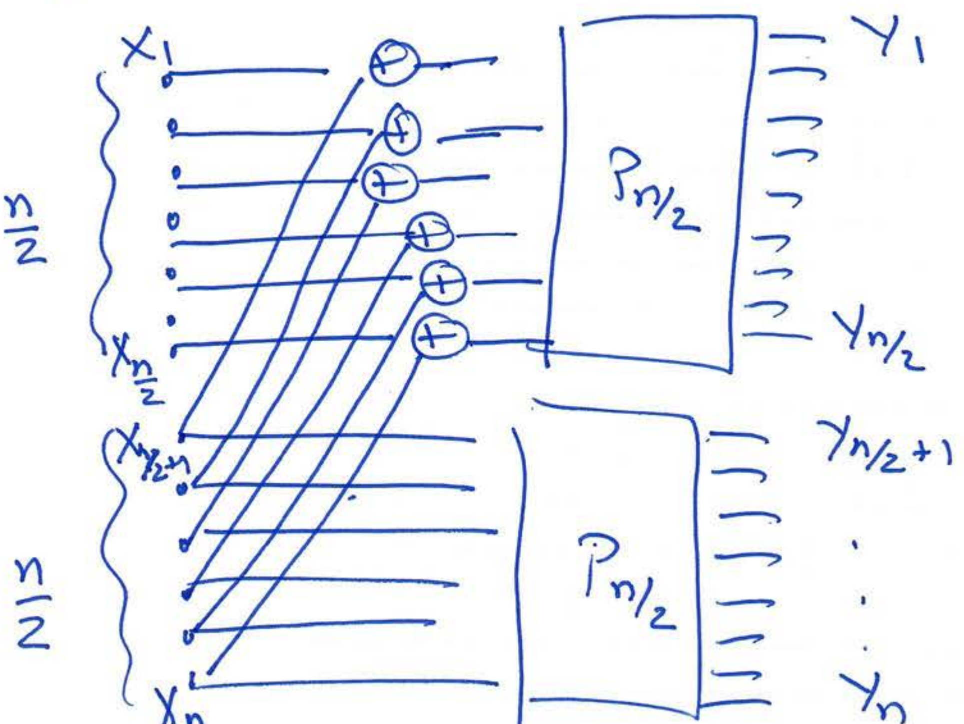
$$① + ② \Rightarrow H(A|C) = H(A|C,D)$$
$$\& \ H(B|D) = H(B|C,D)$$

& so polarization yields

$$H(A \oplus B \mid C,D) > H(A|C,D)$$

$$H(A \oplus B \mid C,D) + H(B \mid A \oplus B, C, D)$$
$$= H(A|C,D) + H(B|C,D).$$

Right n-ary picture



Can verify ① – ④ hold.

Decompression Algorithm

$\left[S \subseteq [N] \right.$ is set s.t.

$\left. \forall\, i \notin S \quad H(Y_i \mid Y_1 \cdots Y_{i-1}) \leq \frac{1}{N^2} \right.$

Task: Given $(Y_i)_{i \in S}$, compute most likely $(Y_i)_{i \notin S}$

given that $Y = P(X)$ & $X \sim \mathrm{Bern}(p)^n$.

~~Outer~~ Algorithm:

Outer loop

for $i = 1 \ldots n$ do

if $i \in S$ do ~~nothing~~. $\hat{Y}_i = Y_i$

if $i \notin S$ compute

$$\boxed{\alpha_i = \Pr_X \left[Y_i = 1 \mid \hat{Y}_1 \cdots \hat{Y}_{i-1} \right]}$$

How? Later!

if $\alpha_i > \frac{1}{2} \Rightarrow Y_i = 1$

else $Y_i = 0$

Analysis: $\Pr_X \left[Y_i \neq \hat{Y}_i \mid (Y_1 \cdots Y_{i-1}) = (\hat{Y}_1 \cdots \hat{Y}_{i-1}) \right] \leq \frac{1}{N^2}$

$H(Y_i \mid Y_1 \cdots Y_{i-1})$

Proof: Probability .... (omitted / exercise) ...

Now compute $q_j = \Pr\left[X_j = 1 \mid X_{j-\frac{n}{2}} \oplus X_j\right]$

(see example)

Recurse on $q_{\frac{n}{2}+1} \dots q_n$ & $Y_{\frac{n}{2}+1} \dots Y_n$.

$\boxed{\text{QED}}$

Clearly takes poly(N) time. Not so clearly takes $O(N \log N)$ time.

$\boxed{\text{QED}}$

$X_1 \backslash X_2$	0	1
0	$(1-p_1)(1-p_2)$	$p_2(1-p_1)$
1	$p_1(1-p_2)$	$p_1 p_2$

$$\Pr\left[\, X_2 = 1 \;\middle|\; X_1 \oplus X_2 = 1 \right] = \frac{p_2(1-p_1)}{p_2(1-p_1) + p_1(1-p_2)}$$

Rest of Lecture

Speed of Polarization

Desired Theorem [STRONG ONE-SIDED]

$\forall\, p$ $\exists$ constant c s.t. $\forall\, \epsilon > 0$

if $N = \Theta\left(\frac{1}{\epsilon}\right)^c$ then P_N has the right # of low-entropy bits :

specific

$$\frac{\#\left\{ i \mid H(Y_i \mid Y_1 \dots Y_{i-1}) \le \frac{1}{N^2} \right\}}{N} \ge 1 - H(p) - \epsilon$$

Theorem: Implied by following two lemmas:

(1) WEAK, ~~ONE~~ TWO - SIDED POLARIZATION

$\exists$ poly s.t. $\forall p, \epsilon > 0$ if $N = \text{poly}(\frac{1}{\epsilon})$

& $(Y_1 \dots Y_n) = P_N(X_1 \dots X_n)$; $X_n \leftarrow \text{Bern}(p)$;

$$\#\{i \mid H(Y_i \mid Y_1 \dots Y_{i-1}) \in (\epsilon, 1-\epsilon)\} \leq \epsilon \cdot N.$$

— x —

what is weak?. Polarized entropies $\leq \frac{1}{N^{.01}}$ or $1 - \frac{1}{N^{.01}}$

& not $\leq \frac{1}{N^2}$ or $1 - \frac{1}{N^2}$

— x —

(2) STRONG ONE-SIDED EXTRA POLARIZATION

$\exists$ poly P_1, P_2, s.t. $\forall \epsilon > 0$ $\boxed{\forall p \leq P_1(\epsilon)}$ $\forall N \geq P_2(\frac{1}{p})$ very small

if $(Y_1 \dots Y_n) = P_N(X_1 \dots X_N)$ & $(X_1 \dots X_n) \leftarrow \text{Bern}(p)^n$

then

$$\#\{i \mid H(Y_i \mid Y_1 \dots Y_{i-1}) \geq \frac{1}{N^3}\} \leq (H(p) + \epsilon) \cdot N$$

One-sided ... ignorable, but conceptual.

1: Proof Steps (mod calculus)

Notation: $p^+ \triangleq 2p(1-p)$

$p^- \triangleq h^{-1}(2h(p) - h(p^+))$

Potential: $\phi(p) \triangleq \sqrt{h(p)(1-h(p))}$

Claim: $\exists \Lambda < 1$ s.t. $\forall 0 < p < \frac{1}{2}$

$$\frac{\phi(p^+) + \phi(p^{-1})}{2} \leq \Lambda(\phi(p))$$

Proof: Calculus, Omitted.

Claim $\Rightarrow$ 1: After ℓ steps of "polarization" $(N = 2^\ell)$

$$\mathop{E}_{i}[\phi(\eta_i)] \leq \Lambda^\ell \quad \text{where } \eta_i = h^{-1}(H(Y_i | Y_1 \ldots Y_{i-1}))$$

$$\Rightarrow \Pr_i\left[\phi(\eta_i) \geq \frac{\epsilon^2}{2^2}\right] \leq \frac{\Lambda^\ell \cdot 4}{\epsilon^2} \leq \epsilon \quad \text{if}$$

$$\ell = \Omega_\Lambda(\log \tfrac{1}{\epsilon})$$

$\Uparrow$

$$h(\eta_i) \in (\epsilon, 1-\epsilon)$$

$\cancel{\Rightarrow \Pr[H(\eta_i) \geq \epsilon]}$

Proof Idea for 2:

Key observation: if p sufficiently small

then $p^+ \leq 2p$

& $p^- \leq \frac{p}{100}$

$\exists\, p_0$ s.t. $\forall\, p \leq p_0$ ⋮

$$\Big[\; h(p) \approx p \log \frac{1}{p}$$

$$h(2p) \approx 2p \log \frac{1}{p} - 2p$$

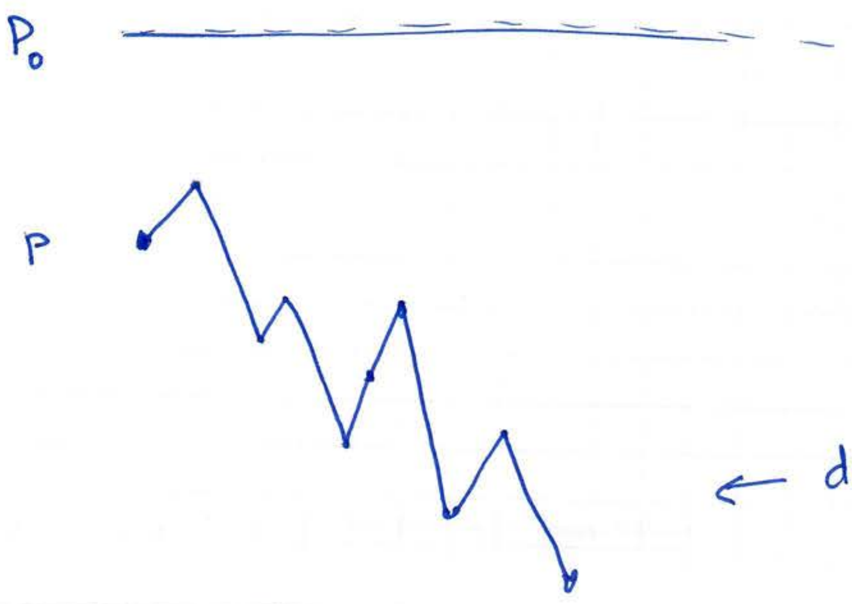
$$2h(p) - h(2p) \approx 2p$$

$$h^{-1}\big(2h(p) - h(2p)\big) \approx \frac{2p}{\log\frac{1}{p}} \;\Big]$$

$\mathbb{E}[\log p]$ decreases by large additive constant. [say 5]

---

m- further steps of polarization starting at small p.



← drift negative!!

$$\Pr_i\left[\text{random walk hits } p_0\right] \approx \cancel{\text{poly}(p_0)}\ \text{poly}(p)$$

$$\Pr_i\left[\cancel{\text{random walk}}\ |\log \eta_i - \log p| < 4m\right] \approx \exp(-m)$$

if neither happens

$$\eta_i \le \frac{p}{\cancel{4}\, 2^{4m}} \le \frac{p}{N^4} \qquad \boxtimes$$

<u>Caveats</u> :- Often dealing with $Y_i \mid Y_0 \cdots Y_{i-1}$ whose Expected entropy is $h(\eta_i)$.

- So need convexity to argue that reasoning about expectations is O.K. ....