Stochastic Coding theory "Correct with high proba randomly chosen to errors"

DEF: $X \in A$ is a random variable on a finite set A.

For $\tilde{x} \in A$. Let $P_X(\tilde{x}) = Pr[X = \tilde{x}]$

$P_X : A \to [0;1]$

If $X \in A$, $Y \in B$. Then $(X, Y) \in A \times B$ is also RV

$P_{XY}(\tilde{x}, \tilde{y}) = Pr[X = \tilde{x}, Y = \tilde{y}]$, $\tilde{x} \in A$, $\tilde{y} \in B$

$P_{Y|X}(\tilde{y}, \tilde{x}) = Pr[Y = \tilde{y} \mid X = \tilde{x}]$

Setting for channel coding

DEF: A channel with inp Alphabet A, output alphabet B is a cond prob distr

$W = P_{Y|X}$, $Y \in B$, $X \in A$. $\quad \forall \tilde{x} \in A \ \forall \tilde{y} \in B \ W(\tilde{y}, \tilde{x}) \geq 0$

$W : A \to B$

$\forall \tilde{x} \in A \ \sum_{\tilde{y} \in B} W(\tilde{y}, \tilde{x}) = 1$

Ex 1: q-ARY Erasure channel

If $|A| = q$ $\quad qEC_\lambda : A \to A \cup \{?\}$ Erasure Prob $\to 0 \leq \lambda \leq 1$

$\forall \tilde{x} \in A$, $qEC_\lambda(? \mid \tilde{x}) = \lambda$

$qEC_\lambda(\tilde{x} \mid \tilde{x}) = 1 - \lambda \Rightarrow qEC_{\forall \tilde{y} \neq \tilde{x}}(\tilde{y} \mid \tilde{x}) = 0$.

Note, if $q = 2$, $2EC_\lambda = BE_\lambda$

Ex 2:

If $|A| = q$ $\quad qSC : A \to A$ $\quad$ q ARY SYMMETRIC CHANNEL ($q = 2$ BSC)

$0 \leq \lambda \leq 1$: err prob

$\forall \tilde{x}$, $qSC(\tilde{x} \mid \tilde{x}) = 1 - \lambda$

$\forall \tilde{y} \neq \tilde{x}$, $qSC(\tilde{y} \mid \tilde{x}) = \frac{\lambda}{q-1}$

Ex 3: $BAWGN_{\sigma^2} : \{+1, -1\} \to \mathbb{R}$

If $X = b \in \{\pm 1\}$, then $Y \sim \mathcal{N}(b, \sigma^2)$

Def: Given a channel $W: A \to B$, $n \geq 1$.

Define **Product Channel** $W^n : A^n \to B^n$

$$W^n(\tilde{y}_1 \dots \tilde{y}_n \mid \tilde{x}_1 \dots \tilde{x}_n) = \prod_{i=1}^{n} W(\tilde{y}_i \mid \tilde{x}_i)$$

Channel Coding: Given $k \leq n$, $W: A \to B$, $\mathcal{C} \subseteq A^n$ $\dim \mathcal{C} = k \mid |\mathcal{C}| = |A|^k$

$$\text{Enc } A^k \to \mathcal{C} \quad \text{(injection)}$$
$$\text{Dec } B^n \to \mathcal{C}$$

Let $M \in A^k$ **The coding process:**

$$M \xrightarrow{A^k} \boxed{\text{Enc}} \xrightarrow{\mathcal{C}} X \to \boxed{W^n} \xrightarrow{B^n} Y \to \boxed{\text{Dec}} \xrightarrow{\mathcal{C}} \hat{X} \to \boxed{\text{Enc}^{-1}} \xrightarrow{A^k} \hat{M}$$

Formally $M \in A^k$ $\quad P_{XY} = P_X P_{Y|X} = P_X W^n \quad \hat{X} = \text{Dec}(Y)$

$X = \text{Enc}(M)$ $\quad \hat{M} = \text{Enc}^{-1}(\hat{X})$

Def: Block error probability

$$P_B[\mathcal{C}, W, \text{Enc}, \text{Dec}] = \max_{\tilde{m} \in A^k} \Pr[\hat{M} \neq M \mid M = \tilde{m}] = \max_{\tilde{x} \in \mathcal{C}} \Pr[\hat{X} \neq \tilde{x} \mid X = \tilde{x}]$$

$P_B(\mathcal{C}, W, \text{Dec})$

Central Problem: Given W, M. Find $\mathcal{C}$, Dec, Enc with

1) Low $P_B(\mathcal{C}, W, \text{Dec})$
2) High $R = k/n$
3) Efficient Dec, Enc, Enc^{-1}

Def Maximum A' posteriori (map) decoder

$\text{DEC}: R^n \to \mathcal{C}$

$$\text{DEC}_{\text{MAP}}(\tilde{y}) = \operatorname*{argmax}_{\tilde{x} \in \mathcal{C}} W^n(\tilde{y} \mid \tilde{x})$$

if max is not unique, choose any

Let $X \sim \text{Unif}(\mathcal{C})$. Then $\text{Dec}_{MAP}(\tilde{y}) = \underset{\tilde{x} \in \mathcal{C}}{\text{argmax}}\, P_{Y|X}(\tilde{y}|\tilde{x}) = \underset{\tilde{x} \in \mathcal{C}}{\text{argmax}} \frac{P_{X|Y}(\tilde{x},\tilde{y})\, P_Y(\tilde{y})}{P_X(\tilde{x})}$

$= \underset{\tilde{x} \in \mathcal{C}}{\text{argmax}}\, P_{X|Y}(\tilde{x},\tilde{y})$

Lemma: Let $\mathcal{C} \subseteq A^n$. $\text{Dec}: B^n \to \mathcal{C}$, $X \sim \text{Unif}(\mathcal{C})$

Then $\Pr[\text{Dec}(Y) \neq X] \geq \Pr[\text{Dec}_{MAP}(Y) \neq X]$

Example: (Map decoder for $q\text{-}\mathcal{E}C_\lambda$)

Let $\mathcal{C} \subseteq A^n$, $\tilde{x} \in \mathcal{C}$, $\tilde{y} \in B^n$, $B = A \cup \{?\}$ then $W^n(\tilde{y}|\tilde{x}) = \begin{cases} 0, & \text{if } \exists i\ \tilde{y}_i \neq ? \wedge \tilde{y}_i \neq x_i \\ \lambda^{wt(\tilde{y},?)} (1-\lambda)^{n-wt(\tilde{y},?)} & \end{cases}$

$\mathcal{C} = \{0000, 0011, 0101, 0110, 1001, 1010, 1100, 1111\}$

$\tilde{y} = 0??1$, then $\tilde{x} \in \{0101, 0011\}$ ($P(\tilde{y}|\tilde{x}) = 0$)

$P(\tilde{y}|\tilde{x}) = \lambda^2$

Map decoder for $q\text{SC}_\lambda$: $\left(\lambda \leq \frac{q-1}{2} \text{ s.t } (1-\lambda) \geq \frac{\lambda}{q-1}\right)$

$$W^n(\tilde{y}|\tilde{x}) = \left(\frac{\lambda}{q-1}\right)^{d_H(\tilde{x},\tilde{y})} (1-\lambda)^{n-d_H(\tilde{x},\tilde{y})}$$

So, $\text{Dec}_{MAP}(\tilde{y}) = \underset{\tilde{x} \in \mathcal{C}}{\text{argmin}}\, d_H(\tilde{x},\tilde{y})$

MAIN QUESTION (ASYMPTOTICALLY)

Given W, what's the largest R s.t $\exists (\mathcal{C}_n, \text{Dec}_n)_n$ $\mathcal{C}_n \subseteq A^n$, $R(\mathcal{C}_n) \to R$

And $P_B(W, \mathcal{C}_n, \text{Dec}_n) \xrightarrow{n \to +\infty} 0$

To answer this qst, we need the notion of Shannon Entropy.

DEF: Let $X \in A$. Then $H(X) = \sum_{\tilde{x} \in A} P(\tilde{x}) \log_2\left(\frac{1}{P(\tilde{x})}\right)$, with $0 \log_2 \frac{1}{0} = 0$.

$P = P_X$

Note: $H(X) = \mathbb{E}\left(\log_2 \frac{1}{P(X)}\right)$.

Entropy measures amount of "Randomness" of "information" in X

Claim: 1) $H(X) \geq 0$ and $H(X)=0$ iff X is constant

2) $H(X) \leq \log|A|$ and $H(X) = \log|A|$ iff $X \sim \text{Unif}(A)$

By Jensen Ineq, φ convex $\Rightarrow \mathbb{E}(\varphi(X)) \geq \varphi(\mathbb{E}(X))$

$\text{supp}(X) = \{\tilde{x} \in A : P_X(\tilde{x}) > 0\}$

Proof of 2): $H(X) = \sum_{\tilde{x} \in \text{Supp}\, A} P(\tilde{x}) \log \frac{1}{P(\tilde{x})} \leq \log\left(\sum P(\tilde{x}) \times \frac{1}{P(\tilde{x})}\right)$ By Jensen

$$\log\left(\sum_{\tilde{x} \in \text{Supp}} 1\right) = \log|\text{Supp}\, X| \leq \log|A| . \blacksquare$$

Conditional Entropy

$X \in A, Y \in B, Z \in C$

$$H(X|Y=\tilde{y}) = \sum_{\tilde{x} \in A} P_{X|Y=\tilde{y}}(\tilde{x}) \log \frac{1}{P_{X|Y=\tilde{y}}(\tilde{x})}$$

$$H(X|Y) = \sum_{\tilde{y} \in B} P_Y(\tilde{y}) H(X|Y=\tilde{y})$$

$$H(X|Y, Z=\tilde{z}) = \sum_{\tilde{y} \in B} P_{Y|Z}(\tilde{y}, \tilde{z}) H(X|Y=\tilde{y}, Z=\tilde{z})$$

Chain Rule of Shannon Entropy: Proof By Simple Computation.

$$H(XY) = H(Y) + H(X|Y) .$$

More Generally: $H(XY|Z) = H(Y|Z) + H(X|YZ)$

More Properties of entropy:

1) If X, Y are independent $(X \perp Y)$

Then $H(XY) = H(X) + H(Y)$

2) Let $f: A \to B$ Then $H(f(X)) \leq H(X)$, Eq when f is injective.
$X \in A$

Proof: As $x \to xf(x)$ is injective, $H(X) = H(Xf(X)) = H(f(X)) + H(X|f(X)) \geq H(f(X))$.

Ex1: $X \sim Ber(p) \Rightarrow H(X) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p} = h(p)$ ← Binary Entropy Fnct

Ex2: $X \sim Unif(\{0,1\})$; $Y =$ $Pr[Y=B \mid X=B] = 1-\varepsilon$

$Pr[Y=1-B \mid X=B] = \varepsilon$

$$H(XY) = H(Y) + H(Y|X) = 1 + \frac{1}{2}\left(H(Y|X=0) + H(Y|X=1)\right) = 1 + h(\varepsilon)$$

Mutual Information:

$$I(X;Y) = H(X) - H(X|Y)$$

Properties:

$$I(X;Y) = I(Y;X) = H(X) + H(Y) - H(XY)$$

Data Processing Inequality:

If $X \perp Z \mid Y$ then $I(X;Z) \leq I(X;Y)$

$X \to Y \to Z$

(Think of it as adding randomness to go from X to Y then more randomness from Y to Z)

Claim: $I(X;Y) \geq 0$ $\left(H(X) \geq H(X|Y)\right)$

Gibbs Inequality:

If $(p_1 \dots p_n)$ $(y_1 \dots y_n)$ prob distributions then

$$\sum_{i=1}^{k} p_i \log \frac{1}{p_i} \leq \sum_{i=1}^{k} p_i \log \frac{1}{q_i}$$

$\Rightarrow$ $\mathbb{E}_{X \sim P} \log \frac{1}{Q(X)}$ is minimized iff $Q = P$, over the set of Prob distr.

$$H(Y) - H(Y|X) = \sum_{\tilde{x},\tilde{y}} P_X(\tilde{x}) P_{Y|X}(\tilde{y},\tilde{x}) \left(\log \frac{1}{P_Y(\tilde{y})} - \log \frac{1}{P_{Y|X}(\tilde{y}|\tilde{x})}\right) = \sum_{\tilde{x},\tilde{y}} P_{XY}(\tilde{x},\tilde{y}) \log \frac{P(\tilde{x},\tilde{y})}{P_Y(\tilde{y}) P_X(\tilde{x})} \geq 0$$

Def: Let $W: A \to B$ be a channel, $|A| = q$

Then $\mathrm{Cap}(W) = \sup_{\substack{XY \in A \times B \\ P_{Y|X} = W}} I(X;Y) \cdot \log_q 2$

(Cap: Capacity)

Example: $W = qEC_\lambda$ $\quad \tilde{x} \xrightarrow{1-\lambda} \tilde{x}$, $\tilde{x} \xrightarrow{\lambda} ?$ $\quad X \to \boxed{W} \to Y$

$$H(X;Y) = H(X) - H(X|Y) = H(X) - \mathbb{P}_r[Y=?]H(X|Y=?) - \sum_{\tilde{x}} \mathbb{P}_r[Y=\tilde{x}]H(X|Y=\tilde{x})$$
$$= H(X) - \lambda H(X) - 0 = (1-\lambda)H(X).$$

$$\mathrm{Cap}(qEC_\lambda) = (1-\lambda)\log q \log_q 2 = (1-\lambda).$$

Example 2,

$\tilde{x} \xrightarrow{1-\lambda} \tilde{x}$

$\tilde{x} \xrightarrow{\frac{\lambda}{q-1}} \tilde{y} \neq \tilde{x}$

$$I(X;Y) = H(Y) - H(Y|X)$$
$$= H(Y) - \sum_{\tilde{x}} P_X(\tilde{x})\, H(Y|X=\tilde{x})$$
$$= H(Y) - \sum_{\tilde{x}} P_X(\tilde{x}) \left((1-\lambda)\log\frac{1}{1-\lambda} + (q-1)\frac{\lambda}{q-1}\log\frac{q-1}{\lambda} \right)$$
$$= H(Y) - \left((1-\lambda)\log_q\frac{1}{1-\lambda} + \lambda \log_q \frac{q-1}{\lambda} \right) \times \log q$$
$$= H(Y) - h_q(\lambda) \times \log q$$

$h_q(\lambda)$: qAry Entropy Function $= (1-\lambda)\log_q\frac{1}{1-\lambda} + \lambda \log_q \frac{q-1}{\lambda}$.

Check: $h_2(\lambda) = h(\lambda)$.

$= \log_2 q (1 - h_q(X))$ (Sup $X \mapsto$ Unif(A), then $Y \mapsto$ Unif(A))

$\Rightarrow \mathrm{Cap}(qSC_\lambda) = 1 - h_q(\lambda)$ WHERE $h_q(\lambda) = (1-\lambda)\log_2\frac{1}{1-\lambda} + \lambda \log_2 \frac{q-1}{\lambda}$

Shannon's Coding Theorem Let $W : A \to B$

1/ $\forall R > Cap(W)$, $\exists M_0$, $\forall m \geq M_0$ $\underset{R(\mathcal{C}) \geq R}{\mathcal{C} \subseteq A^m}$ $Dec : B^m \to \mathcal{C}$

$$P_B(\mathcal{C}, W, Dec) \geq 0.9 \quad \text{(or any constant)}$$

2) $\forall R < Cap(W)$ $\exists \underset{(Dec_n)_n}{(\mathcal{C}_n)_n}$ $\mathcal{C}_n \subseteq A^n$, $R(\mathcal{C}_n) \geq R$

$$\lim_{n \to +\infty} P_B(\mathcal{C}_n, W, Dec_n) = 0$$

PROOF OF WEAK NEGATIVE

For $W : A \to B$, $\forall R > Cap(W)$, $\exists c > 0$

$\forall n \in \mathbb{N}$, $\forall \underset{Dec : B^n \to \mathcal{C}}{\mathcal{C} \subseteq A^n}$ s.t. $R(\mathcal{C}) \geq R$ Then $P_B(\mathcal{C}, W, Dec) \geq c$

Fano Inequality:

Let $X \in A$, $Y \in B$, $f : B \to A$ $\hat{X} = f(Y)$

$P = P_r[X \neq \hat{X}]$

Then:

$$H(X|Y) \leq h(p) + p \log(|A| - 1)$$

Proof of Fano's

$$H(X|Y) = H(X|\hat{X}Y) \leq H(X|\hat{X})$$

Let $E = \mathbb{1}[X \neq \hat{X}]$

$$H(X|\hat{X}) = H(EX|\hat{X}) = H(E|\hat{X}) + H(X|E\hat{X}) \leq H(E) + P(E=0)\underbrace{H(X|\hat{X}, X=\hat{X})}_{0} + P(E=1) H(X|\hat{X}, X \neq \hat{X})$$

$$\leq h(p) + p \log(|A| - 1). \blacksquare$$

$W : A \to B$ $\quad Cap(W) = \max_{X,Y : P_{Y|X} = W} I(X;Y) \cdot \log_q 2$ $\quad | \quad X \to \boxed{W} \to Y$, then $I(X;Y) \leq \log q \cdot Cap(W)$

$|A| = q$

FANO $X \in A, Y \in B$, $f : B \to A$, $\hat{X} = f(X)$, $p = Pr(\hat{X} \neq X)$

$$\Downarrow$$

$$H(Y|Y) \leq h(p) + p \log(|A| - 1)$$

Negative Shannon: $W : A \to B$, $R > Cap(W)$, $\exists \varepsilon > 0$ s.t. $\forall \mathcal{C} \subseteq A^n$, $R(\mathcal{C}) \geq R$, $Dec : B^n \to \mathcal{C}$

$$P_B(\mathcal{C}, W, Dec) \geq \varepsilon$$

Proof: $Cap(W^n) = n \, Cap(W)$

Let P_X s.t. $I(X;Y) = \log q \cdot Cap(W)$ Then if $X_1^n \sim (P_X)^n \Rightarrow I(X_1^n ; Y_1^n) = n \, I(X_1 ; Y_1) = n \log q \, C$

$$= \sum_{i=1}^{n} I(X_1, Y_1)$$

$$\Downarrow$$

$$Cap(W^n) \geq n \, Cap(W)$$

$$\leq \leftarrow \text{By Induction}$$

$X \sim Unif(\mathcal{C})$, $Dec : B^n \to \mathcal{C}$

Fano $X, Y, \hat{X} = Dec(Y)$, $p = Pr[Dec(Y) \neq X]$

$$H(X|Y) \leq h(p) + p \log(|A| - 1) \leq 1 + pn \log q$$

$$H(X) - I(X;Y) \geq \log q^{Rn} - \log q \cdot n \cdot Cap(W) = n \underbrace{\log q (R - Cap(W))}_{\varepsilon > 0} = n\varepsilon$$

where $H(X) = Rn \log q$.

$$\varepsilon n \leq 1 + pn \log q \Rightarrow p \geq \varepsilon' > 0 \quad \boxed{\text{for } n \geq n_0}$$

Positive Shannon:

$W: A \to B$ $R \in \mathcal{C}(W) \Rightarrow \exists (\mathcal{C}_n)_n$ $R(\mathcal{C}_n) \geq R$ s.t $\lim_{n\to\infty} P_B(\mathcal{C}_n, W, Dec_{MAP}) = 0$

Let P_X achieve $I(X;Y) = Cap(W) \cdot \log q$

$$Z(\tilde{x},\tilde{y}) = \log \frac{P_{XY}(\tilde{x},\tilde{y})}{P_X(\tilde{x})P_Y(\tilde{y})} \qquad \tilde{x}\in A, \tilde{y}\in B$$

1) $\mathbb{E}\, Z(\tilde{x},\tilde{y}) = \sum_{\tilde{x},\tilde{y}} P_{XY}(\tilde{x},\tilde{y}) \log \frac{P_{XY}(\tilde{x},\tilde{y})}{P_X(\tilde{x})P_Y(\tilde{y})} = H(X)+H(Y)-H(XY) = I(X,Y) = Cap(W) \times \log q.$

2) For $\tilde{y}\in B$ $\quad \underset{\tilde{x}\in A}{\operatorname{argmax}}\, Z(\tilde{x},\tilde{y}) = \underset{\tilde{x}\in A}{\operatorname{argmax}}\, P_{Y|X}(\tilde{y}|\tilde{x}) = Dec_{MAP}(\tilde{y})$

Consider $X_1 \dots X_n$ $\quad \overline{X_1}, \dots, \overline{X_n}$ $\quad (X_i, \overline{X_i})$ iid $\sim P_X$

$X_i \perp \overline{X_i}$

$X_1 \to [W] \to Y_1$ $\quad X_n \to [W] \to Y_n$

3) $Pr\left[\left| \sum_{i=1}^{n} Z(X_i;Y_i) - \log q \cdot Cap(W)\, n \right| > \delta_n \right] \xrightarrow{n\to\infty} 0$ $\quad$ (WLLN)

$\delta > 0$

4) $Pr\left[\sum_{i=1}^{n} Z(\overline{X_i};Y_i) \geq \tau \right] \leq 2^{-\tau}$

$\tau > 0$

$R < Cap(W)$, take $M = q^{\lceil Rn \rceil + 1}$, $\tau = \log q \frac{Cap(W)-R}{2} M$

Choose $\mathcal{C}$ randomly: $\mathcal{C} = (\tilde{X}^{(1)}, \dots, \tilde{X}^{(M)})$ $\quad \tilde{X}^{(j)} \sim P_X$ iid

Let $J \sim Unif(\{1 \dots M\})$

$\tilde{X}^{(J)} = X \to [W^n] \to Y \to [Dec] \to \hat{J}$ $\quad Dec(\tilde{y}) = \underset{j}{\operatorname{argmax}} \sum_{i=1}^{n} Z(\tilde{X}_i^{(j)}, \tilde{y}_i)$

$$Pr[\hat{J} \neq J \mid J = 1] \leq Pr\left[\sum_{i=1}^{n} Z(\tilde{X}_i^{(1)}, Y_i) \leq \log M + \delta\right] + Pr\left[\exists_{j>1} : \sum_{i=1}^{n} Z(\tilde{X}_i, Y_i) > \log M + \delta\right]$$

CHERNOFF/HOEFFDING BOUNDS $\Rightarrow$ EXP. SMALL LARGE DEVIATIONS

Repeating. $\forall j \Pr_{J,\mathcal{E},Y}[J \neq \hat{J}] \xrightarrow{n\to\infty} 0$

$\exists (\mathcal{C}_n)_n \Pr_{J,Y}[J \neq \hat{J}] \to 0$

WANT TO SHOW

$P_B \to 0$

$\| \max_j \Pr[\hat{J} \neq J \mid J = j]$

Assume that $\Pr_{J,Y}[J \neq \hat{J}] \leq \varepsilon$

Let $P_j = \Pr[J \neq \hat{J} \mid J = j]$

$\#BAD = |\{j : P_j \geq 2\varepsilon\}|$

Then $\varepsilon \geq \Pr[\hat{J} = J] = \frac{1}{2^n} \sum_{j=1}^{M} P_j$ $\overset{\text{Markov}}{\Rightarrow}$ $\#BAD \leq \frac{M}{2}$

After discarding $\leq \frac{M}{2}$ elements

we get a code with $P_E \leq 2\varepsilon$. ∎

- Can be also proved with $\mathcal{E}P$.
- Combinatorial proofs for BEC / BSC

  SEE Venkat's book.
- For "Additive" channels including qEC / qSC, random linear codes also achieve capacity.

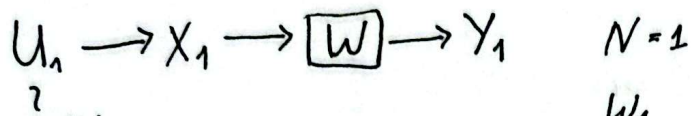
A family $(\mathcal{C}_n)_n$ with $R(\mathcal{C}_n) \to Cap(W)$

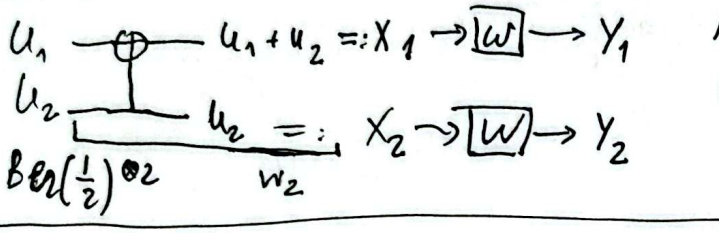
$P_E(\mathcal{C}_n, W, Dec_n) \to 0$

is said to achieve capacity on W using PPC_M.

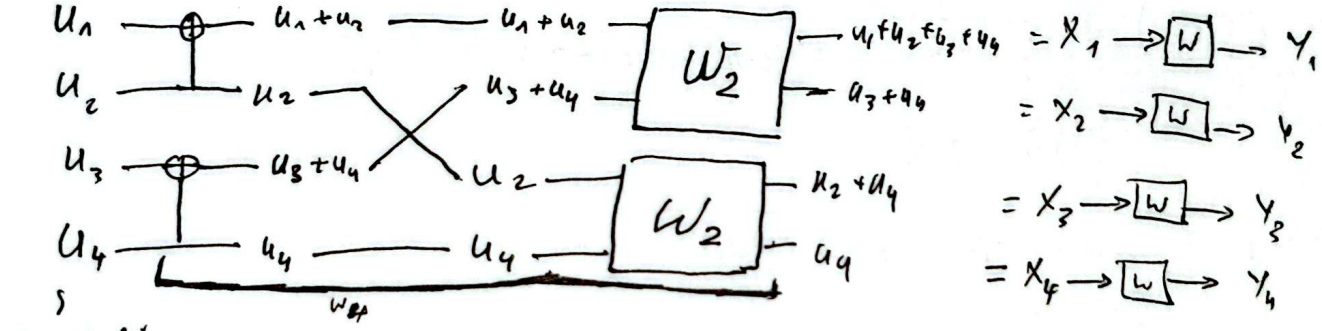
Polar Codes (How to achieve capacity efficiently)

$W : \mathbb{F}_2 \to B$ | Polar Transform

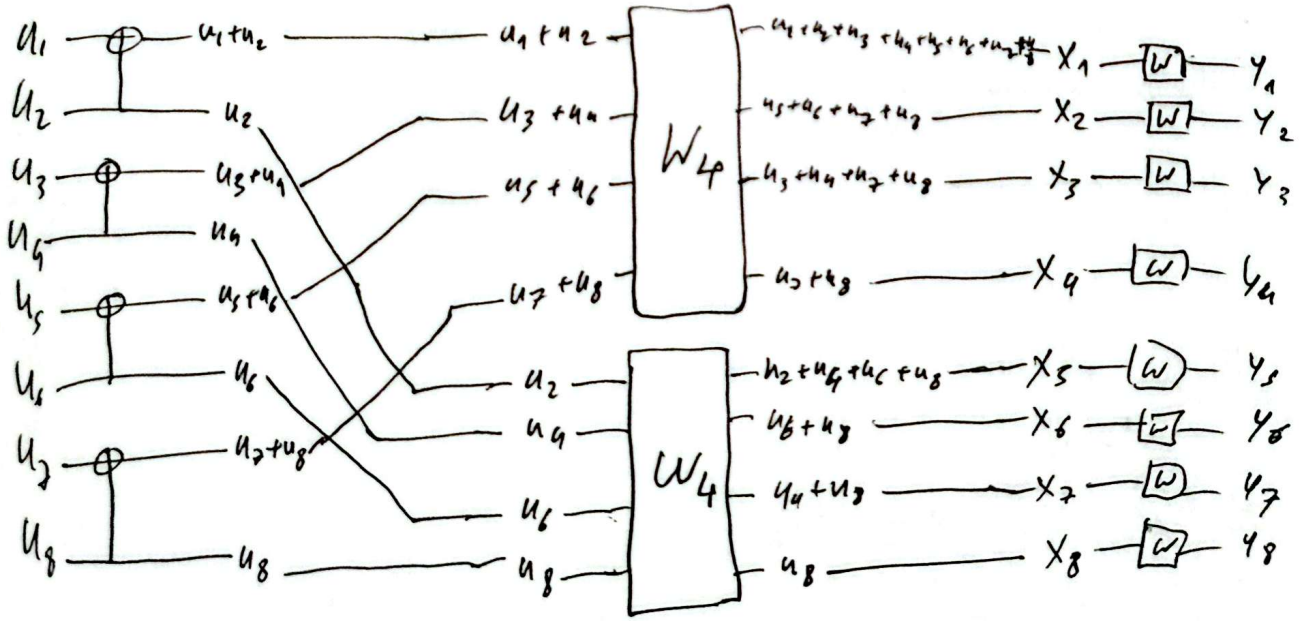




$N = 4$, W_4



$N = 8$, W_8



Def: Let $W : \mathbb{F}_2 \to B$

$$Z(W) = \sum_{\hat{y} \in B} \sqrt{W(\hat{y}|0) W(\hat{y}|1)} \leftarrow \text{Battacharyya coefficient}$$

$I(W) = I(X;Y)$ where $X \sim \text{Ber}(1/2)$, $P_{Y|X} = W$

(W particular : if W achieves capacity for $X \sim \text{Ber}(1/2)$ then $I(W) = \text{Cap}(W)$.

If $I(W) = 0$, then $W(y|0) = W(\hat{y}|1) \ \forall \hat{y} \Rightarrow Z(W) = \sum \sqrt{W(y|0) W(y|1)} = \sum W(y|0) = 1$.

Hence, good channel $\Leftrightarrow Z(W)$ is low.

Example :

$$Z(BEC_\lambda) = \sqrt{W(0|0) W(0|1)} + \sqrt{W(1|0) W(1|1)} + \sqrt{W(?|0) W(?|1)}$$

$$= \sqrt{\lambda \cdot 1} + \lambda.$$

$$Z(BSC_\lambda) = \sqrt{W(0|0) W(0|1)} + \sqrt{W(1|0) W(1|1)}$$

$$= 2\sqrt{\lambda(1-\lambda)}.$$

Properties :

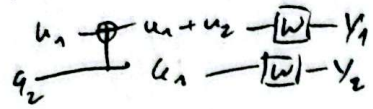
- $I(W) \geq \log \dfrac{2}{1 + Z(W)}$
- $I(W) \leq \sqrt{1 - Z(W)^2}$

Polar Channels : For $N = 2^n$, $\forall i \in [1; n]$

Define a channel $W_N^{(i)} : \mathbb{F}_2 \to B^n \times \mathbb{F}_2^{i-1}$

$$W_N^{(i)}(\tilde{y}_1 \cdots \tilde{y}_n \tilde{u}_1 \cdots \tilde{u}_{i-1} | \tilde{u}_i) \ P_{Y_1 \cdots Y_n U_1 \cdots U_{i-1} | U_i}(\tilde{y}_1 \cdots \tilde{y}_n \tilde{u}_1 \cdots \tilde{u}_{i-1} | \tilde{u}_i)$$

$N=2$:



$W_2^{(1)} : \mathbb{F}_2 \to B^2$

$$W_2^{(1)}(y_1y_2|u_1) = P_{Y_1Y_2|U_1}(y_1y_2|u_2) = \frac{1}{2}\left(P(y_1y_2|u_1u_2=0) + P(y_1y_2|u_1,u_2=1)\right)$$

$$= \frac{1}{2}\left(W(y_1|u_1)W(y_2|0) + W(y_1|u_1\oplus 1)W(y_2|1)\right)$$

$$W_2^{(2)}(y_1y_2u_1|u_2) = \frac{1}{2}W(y_1|u_1\oplus u_2)W(y_2|u_1)$$

$$Z(W_2^{(2)}) = \sum_{y_1y_2u_1}\sqrt{W^{(2)}(y_1y_2u_1|0)W^{(2)}(y_1y_2u_1|1)}$$

$$= \sum_{y_1y_2u_1}\frac{1}{2}\sqrt{W(y_2|u_1)W(y_2|0)W(y_1|u_1\oplus 1)W(y_2|1)}$$

$$= \sum_{y_1y_2u_1}\frac{1}{2}\sqrt{W(y_1|0)W(y_2|0)W(y_1|1)W(y_2|1)}$$

$$= \sum_{y_1y_2}\sqrt{W(y_1|0)W(y_2|0)W(y_1|1)W(y_2|1)}$$

$$= \left(\sum_{y_1}\sqrt{W(y_1|0)W(y_1|1)}\right)\left(\sum_{y_2}\sqrt{W(y_2|0)W(y_2|1)}\right) = Z(W)^2$$

So $W_2^{(2)}$ is "~~less~~ noisy" than W.

On the other hand, $Z(W_2^{(1)}) \geq Z(W)$

Hence by iterating N, some channels become better, and others worse.

Prop: $I(W_{2N}^{(2i)}) + I(W_{2N}^{(2i-2)}) = 2I(W_N^{(i)})$

$Z(W_{2N}^{(2i)}) = Z(W_N^{(i)})^2$

$Z(W_{2N}^{(2i-1)}) \geq Z(W_N^{(i)})$

$W_2^{(1)} = P_{Y_1Y_2|U_1}$

$W_2^{(2)} = P_{Y_1Y_2U_1|U_2}$

RECALL: $I(W) = I(X;Y) = I(U;Y)$ $\quad U - X - \boxed{W} - Y$

$I(W_2^{(1)}) = I(U_1; Y_1Y_2)$

$I(W_2^{(2)}) = I(U_2; Y_1Y_2U_1)$

$I(W_2^{(1)}) + I(W_2^{(2)}) = H(U_1) - H(U_1|Y_1Y_2) + H(U_2) - H(U_2|U_1Y_1Y_2)$

$= 2 - H(U_1U_2|Y_1Y_2) \underset{\text{chain rule}}{=} 2H(X_1X_2|Y_1Y_2)$

$= 1 - H(X_1|Y_1) + 1 - H(X_2|Y_2)$

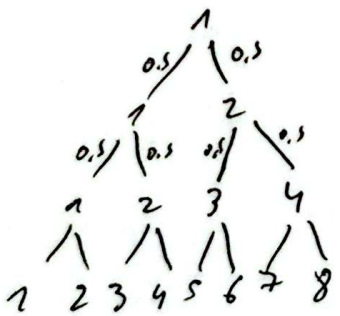
$= H(X_1) - H(X_1|Y_1) + H(X_2) - H(X_2|Y_2)$

$= 2I(W).$

Consider Random Process

$J_0 = 1$. $\forall n$ $P_r(J_{n+1} = 2i \mid J_n = i] = \frac{1}{2}$

$P_r(J_{n+1} = 2i-1 \mid J_n = i] = \frac{1}{2}$



Let $W(n) = W_{2^n}^{J(n)}$

$Z(n) = Z(W(n))$

$I(n) = I(W(n))$

Then $E[I(n+1) \mid I(n)] = \frac{1}{2}\left(I(W_{2N}^{2i}) + I(W_N^{(2i+1)})\right)$

$= I(W_n^{(i)}) = I(n)$

A Random Process $(X(n))_n$ which satisfies

$\mathbb{E}[X(n+1) \mid X(n)] = X(n)$ is a martingale

$\mathbb{E}[X(n+1) \mid X(n)] \leq X(n)$ is a supermartingale

THM (Martingale convergence)

Let $I(0), \ldots, I(n), \ldots$

Be supermartingale satisfying $|I(n)| \leq M \ \forall n$

Then $\exists$ (Almost surely) $I(\infty)$ s.t

$$I(\infty) = \lim_{n \to \infty} I(n)$$

In particular $(I(n))_n$ $(Z(n))_n$ converge almost surely

$(Z_n)_n$ is supermartingale. (no proof)

Furthermore,

$$\text{Supp}(Z_\infty) = \text{supp}(I(\infty)) \in \{0,1\}$$

$$\downarrow$$

$$\mathbb{E}(I(\infty)) = \mathbb{E}[I(0)] = I(W)$$

So $I(\infty) \sim \text{Ber}(I(W))$

In the limit $\approx I(W) \cdot N$ channels satisfies $I(W_N^{(i)}) \approx 1$

$$- H(U_i \mid Y_1^N U_1^{i-1}) \simeq 0$$

$\approx (1 - I(W)) \cdot N$ channels satisfies $I(W_N^{(i)}) \simeq 0$

Given N, ε let $S(N,\varepsilon) = \{i : I(W_N^{(i)}) \geq 1-\varepsilon\}$

Let $K = |S(N,\varepsilon)|$

To encode K bits of information $z_1 \dots z_K$ put them in $(u_i)_{i \in S}$

then set $(u_i)_{i \notin S} = 0$

Apply polar transform.

To decode, use map decoding for $W_N^{(i)}$

$$\underset{w_i}{\operatorname{argmax}}\, W(y_1 - y_N u_1 - u_{i-1} | w_i)$$

Proceed successively on i.