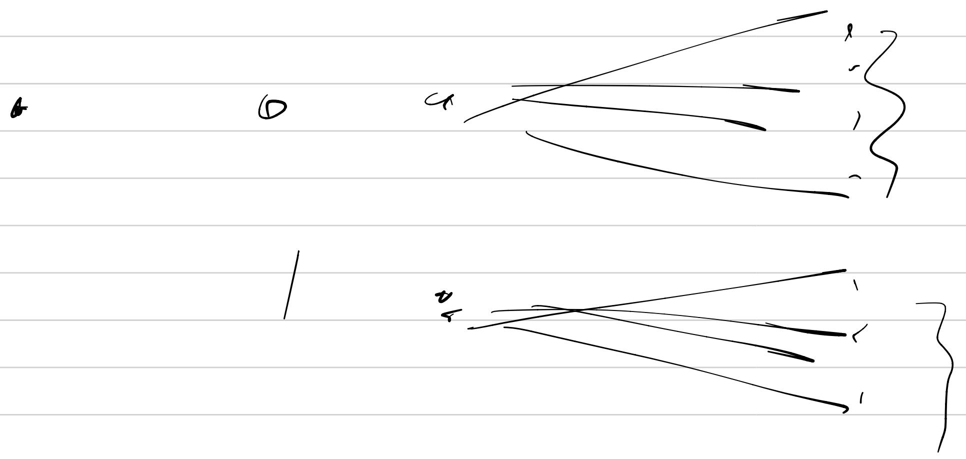
Information Theory & Coding

Nov 30, 2020

Today: "Polar Coding": A coding scheme for Binary input channels (output need not be binary) with the following properties:

- all rates up to $I(X;Y)\big|_{X\sim\text{unif}\{0,1\}}$ is supported
- enc & dec functions can be implemented with low computational complexity.
    - $O(n\log n)$ operations (n = block length)
- Prob. error $\approx 2^{-\sqrt{n}}$.
- the identification of enc & dec is explicit, and can be done in low complexity.

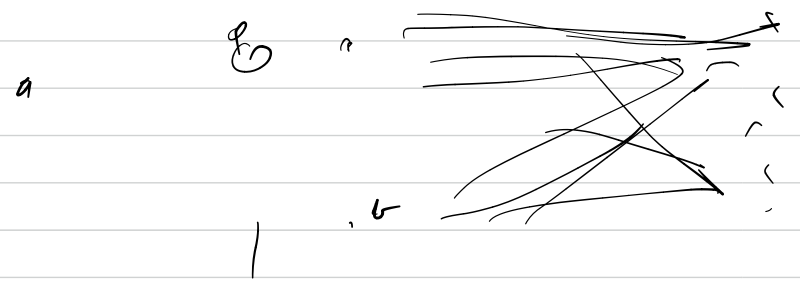
Examples of "easy" channels. (binary input)



$p(y|0)\cdot p(y|1) = 0$

$\forall y$.

$C = 1$, trivially achieved



$p(y|0) \equiv p(y|1)$

$\equiv$ Y is indep of X.

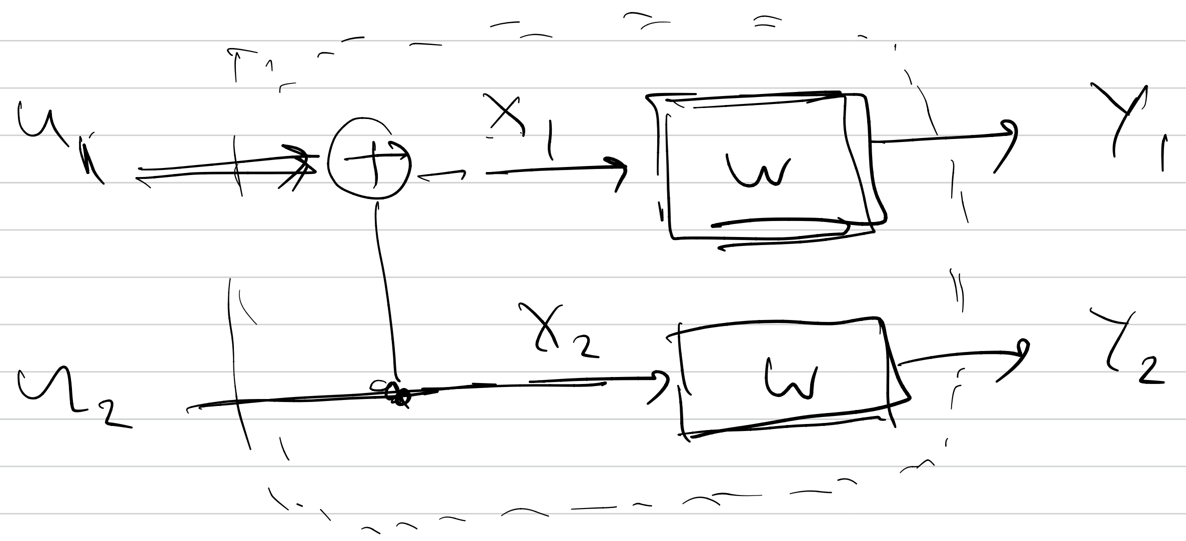
$C = 0$, trivially ach.

The "polar coding" works by constructing such extremal channels from a given "normal" channel;

This is done by a procedure I will call "polar transform".

Suppose we have a channel (Disc., Mem.).

$$W : \{0,1\} \longmapsto \mathcal{Y},$$



$$X_1 = U_1 \oplus U_2 \qquad \Longleftrightarrow \qquad U_1 = X_1 \oplus X_2$$
$$X_2 = U_2 \qquad\qquad\qquad\qquad U_2 = X_2$$

Suppose $\underline{U_1, U_2}$ are indep & $B(\frac{1}{2})$

$\equiv$ $(U_1 U_2)$ uniform on $\{0,1\}^2$

$\equiv$ (X_1, X_2) " " "

$\equiv$ X_1 X_2 are indep & $B(\frac{1}{2})$

$$I(U_1 U_2 ; Y_1 Y_2) = I(U_1 ; Y_1 Y_2) + \underline{I(U_2 ; Y_1 Y_2 \mid U_1)}$$
$$= I(U_1 ; Y_1 Y_2) + \underline{I(U_2 ; Y_1 Y_2 U_1)}$$

$$I(U_1 U_2; Y_1 Y_2) = I(U_1; Y_1 Y_2) + I(U_2; Y_1 Y_2 U_1)$$

$$\|$$

$$I(X_1 X_2; Y_1 Y_2) = I(X_1; Y_1) + I(X_2; Y_2)$$

indep of X_1, X_2

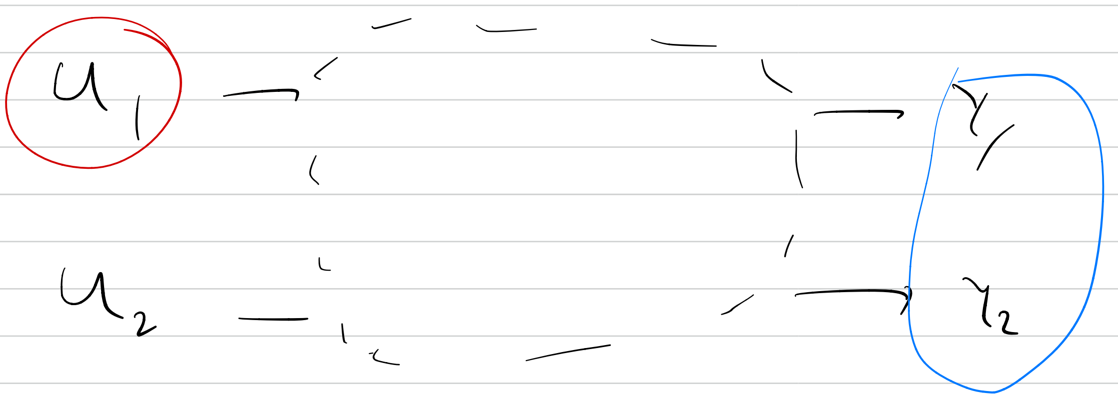
& memoryless channel

$$= 2 I(W)$$

"$I(\text{channel}) = I(\text{input}; \text{output}) \big|_{\text{input are uniform}}$"

So far:

$$2 I(W) = I(U_1; Y_1 Y_2) + I(U_2; Y_1 Y_2 U_1).$$



$$= I(W^-) + I(W^+).$$

$$W^- : U_1 \longmapsto Y_1 Y_2$$

$$W^-(y_1 y_2 \mid u_1) = \sum_{u_2 \in \{0,1\}} \frac{1}{2} W(y_1 \mid u_1 + u_2) W(y_2 \mid u_2)$$

$$W^-(y_1 y_2 | 0) = \frac{1}{2} W(y_1|0) W(y_2|0) + \frac{1}{2} W(y_1|1) W(y_2|1)$$

$$W^-(y_1 y_2 | 1) = \frac{1}{2} W(y_1|0) W(y_2|1) + \frac{1}{2} W(y_1|1) W(y_2|0).$$

$$W^+ : U_2 \to Y_1 Y_2 U_1$$

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

$$W^+(y_1 y_2 0 | 0) = \frac{1}{2} W(y_1|0) W(y_2|0)$$

$$W^+(y_1 y_2 1 | 0) = \frac{1}{2} W(y_1|1) W(y_2|0)$$

$$W^+(y_1 y_2 0 | 1) = \ldots$$

$$W^+(y_1 y_2 1 | 1) = \ldots$$

So we have constructed synthetic channels W^+ & W^- from two uses of the "real" channel W.

"Diversion": in what sense is W^+ a channel?

The job of the receiver is to reconstruct inputs $\underline{(U_1 U_2 \cdots U_n)}$ from the observation $Z = (Y_1 \cdots Y_n)$. Consider the following

"$\underline{\text{magic}}$" method.

$$\hat{U}_1 = \phi_1(Z) \qquad U_1 \to Z$$

$$\hat{U}_2 = \phi_2(Z U_1) \qquad U_2 \to Z U_1$$

$$\hat{U}_3 = \phi_3(Z U_1 U_2)$$

$$\vdots$$

$$\hat{U}_n = \phi_n(Z U^{n-1}) \qquad U_n \to Z U^{n-1}$$

channels which govern the magic method

We can now consider a non-magic method, based on the same decoding functions ϕ_i:

$$\tilde{U}_1 = \phi_1(Z)$$

$$\tilde{U}_2 = \phi_2(Z \tilde{U}_1)$$

$$\tilde{U}_3 = \phi_3(Z \tilde{U}_1 \tilde{U}_2)$$

$$\vdots$$

$$\tilde{U}_n = \phi_n(Z \tilde{U}_1^{n-1}).$$

observe that if (u^n, z) is such that

$\hat{u}^n = u^n$ (ie, if the magic decoder is correct)

then $\tilde{u}_1 = \phi_1(z) = \hat{u}_1 = u_1$

$\tilde{u}_2 = \phi_1(z, u_1) = \hat{u}_2 = u_2$

$\vdots$

$\tilde{u}_n = \phi_n(z u^{n-1}) = \hat{u}_n = u_n$

$\Rightarrow \tilde{u}^n = u^n$.

$\Rightarrow \Pr(\tilde{u}^n \neq u^n) \le \Pr(\hat{u}^n \neq u^n)$

$=$

So for the purpose of analyzing the non-magic system we can analyze the magic system instead.

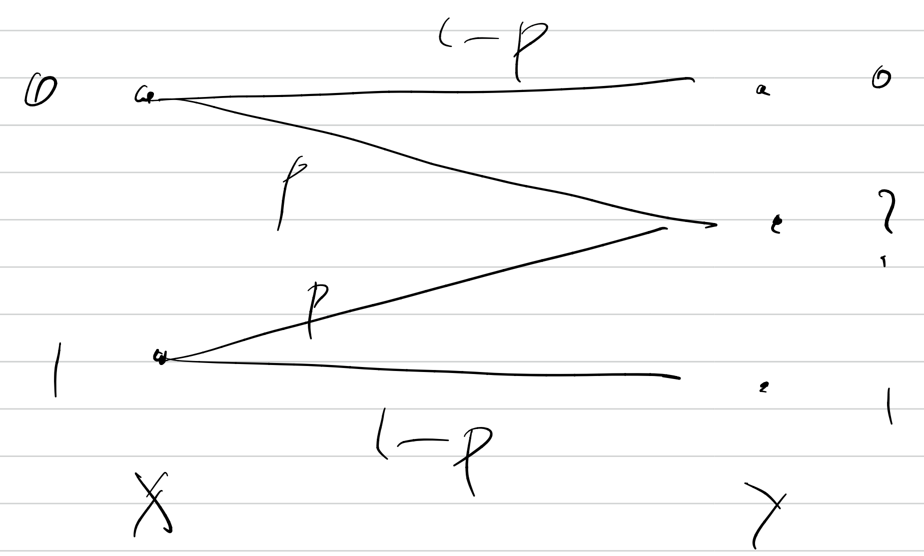
Consequently we can analyze the channels

$W^-: u_1 \to Y_1 Y_2$ & $W^+: u_2 \to Y_1 Y_2 u_1$

(instead of $W^-: u_1 \to Y_1 Y_2$; $W^+: u_2 \to Y_1 Y_2 \tilde{u}_1$)

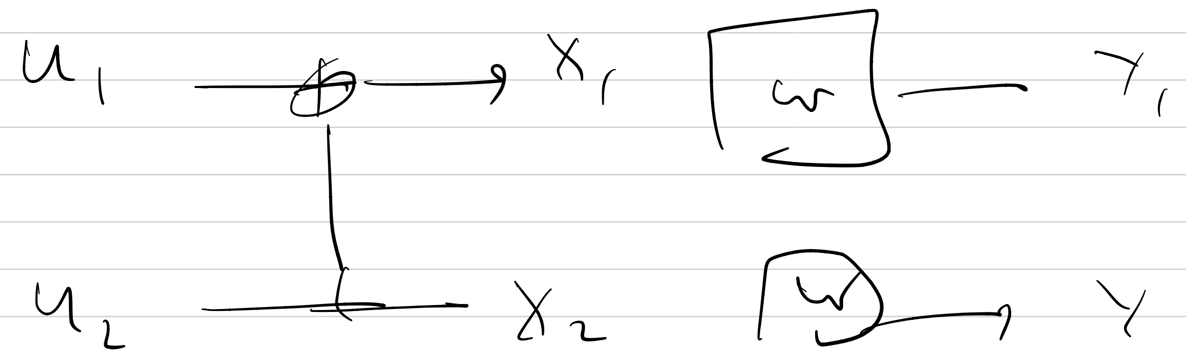
to compute the error probability of the receiver.

Example: $W = BEC(p)$



$$Y = \begin{cases} X & \text{w.p. } 1-p \\ ? & \text{w.p. } p \end{cases}$$

$I(W) = 1-p.$



$$Y_1 Y_2 = \begin{cases} X_1 X_2 & \text{w.p. } (1-p)^2 \\ X_1 ? & \text{w.p. } p(1-p) \\ ? X_2 & \text{"} \\ ? ? & p^2 \end{cases}$$

$$W^- : u_1 \to Y_1 Y_2 = \begin{cases} u_1+u_2,\ u_2 & \text{w.p. } (1-p)^2 & \equiv u_1 \\ u_1+u_2,\ ? & \text{w.p. } (1-p)p & \equiv ? \\ ?\ \ u_2 & \text{" "} & \equiv ? \\ ?\ \ ? & p^2 & \equiv ? \end{cases}$$

$$W^- \equiv BEC(p^2 + 2p(1-p)) = BEC(p(2-p)).$$

$$W^+ : U_2 \to Y_1 Y_2 U_1 = \begin{cases} U_1+U_2, U_2, U_1 & \text{w.p. } (1-p)^2 \\ \boxed{U_1+U_2, ?, U_1} & \text{w.p. } p(1-p) \\ ?\ \ U_2, U_1 & \text{''} \\ ?\ \ ?\ \ U_1 & p^2 \end{cases}$$

$$\equiv \begin{cases} U_2 & \text{''} \\ U_2 & \text{''} \\ U_2 & \\ ? & p^2 \end{cases}$$

$$\Rightarrow W^+ \equiv BEC(p^2).$$

$$W = BEC(p) \Rightarrow W^- = BEC(2p-p^2)$$
$$W^+ = BEC(p^2)$$

$$2I(W) = I(W^-) + I(W^+)$$
$$2(1-p) = (1-(2p-p^2)) + (1-p^2)$$

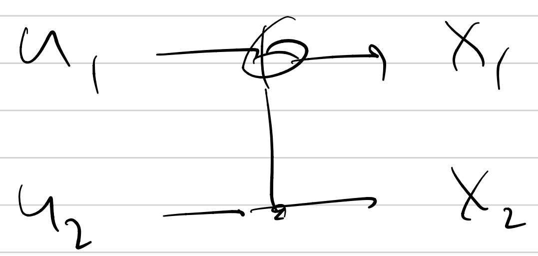
since $p^2 \le p \Rightarrow I(W^+) \geq I(W) \geq I(W^-)$

Observe:

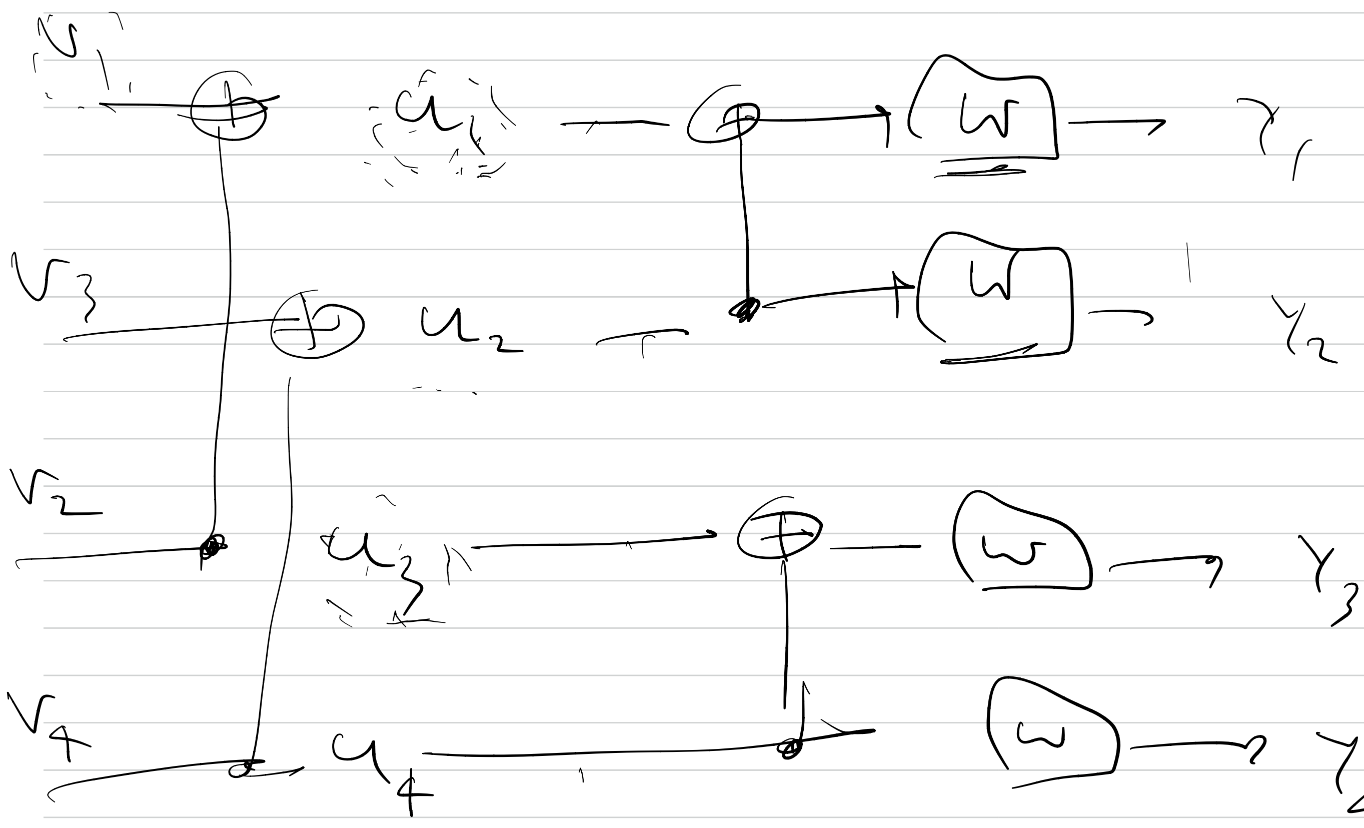
$$I(W^+) = I(U_2; Y_1 Y_2 U_1) \geq I(U_2; Y_2)$$
$$= I(X_2; Y_2) = I(W)$$

So $I(W^+) \geq I(W) \geq I(W^-)$ is true in general.

Building block of the polar transform is



The idea is to iterate:



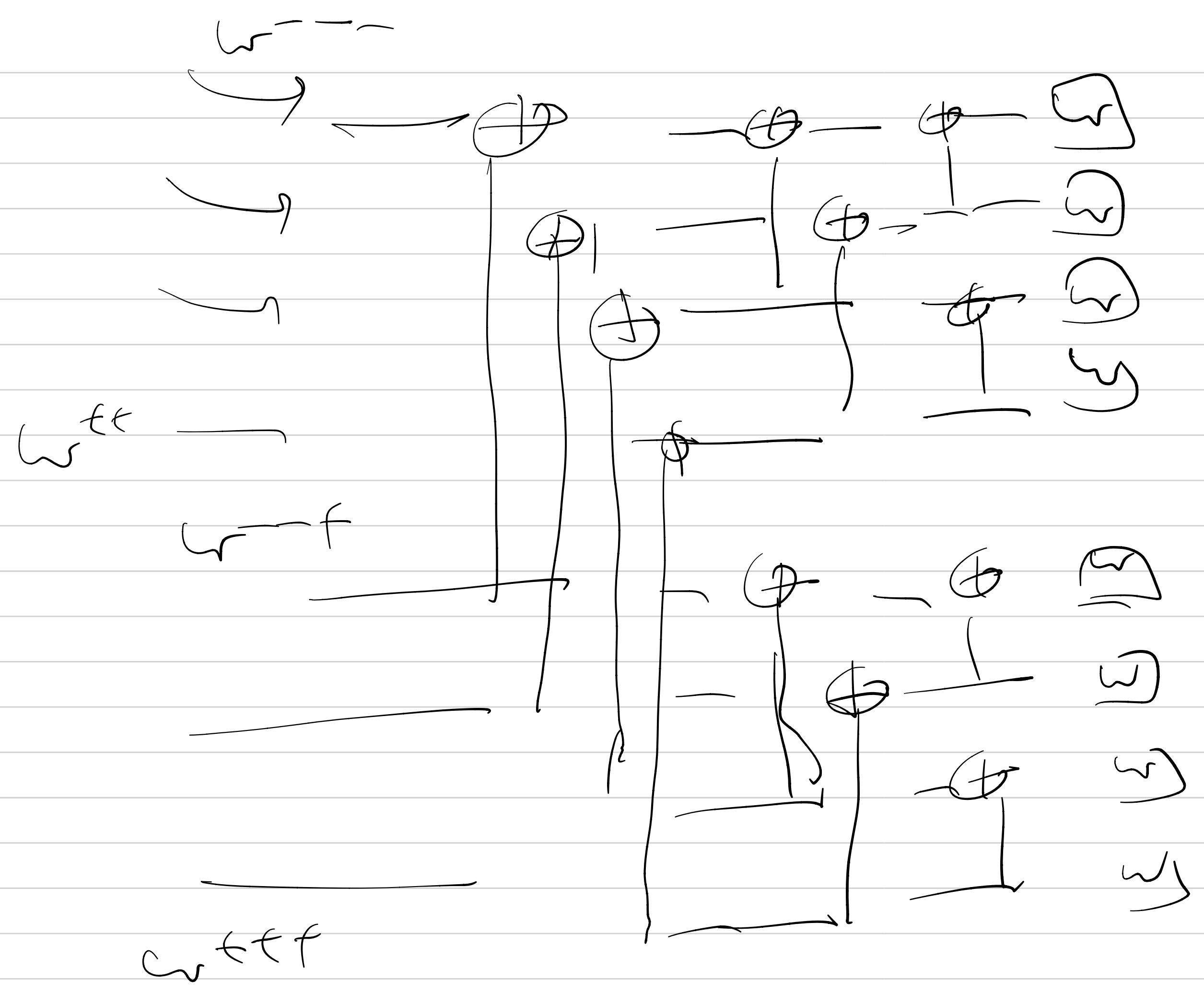
W^- : $u_1 \to Y_1 Y_2$

W^+ : $u_2 \to Y_1 Y_2 u_1$

W^- : $u_3 \to Y_3 Y_4$

W^+ : u_4 : $Y_3 Y_4 u_3$

- W^{--} : $V_1 \to (Y_1 Y_2)(Y_3 Y_4)$
- W^{-+} : $V_2 \to Y_1 Y_2 Y_3 Y_4 V_1$
- W^{+-} : $V_3 \to Y_1 Y_2 u_1 \; Y_3 Y_4 u_3 \equiv Y_1 Y_2 Y_3 Y_4 V_1 V_2$
- W^{++} : V_4 : $Y_1 Y_2 u_4 \; Y_3 Y_4 u_3 V_3 \equiv Y_1 Y_2 Y_3 Y_4 V_1 V_2 V_3$.



By iterating this t times we will have:

use the "real" channel W 2^t times to

obtain one copy each of the channels

$$\underbrace{W^{\overbrace{--\cdots-}^{t}}, \ldots, W^{\overbrace{++\cdots+}^{t}}}$$

$$\equiv \{ W^s ; s \in \{+,-\}^t \} \quad : \text{synthetic channels}$$

We hope that these synthetic channels will be extremal.

Numerical experiment we performed suggests the following theorem:

Thm: With $W = BEC(p)$, we know that

W^s is also a $BEC(p^s)$. We claim that $\forall \varepsilon > 0$

$$\frac{1}{2^t} \sum_{s \in \{+,-\}^t} \mathbb{1}\{ p^s \in (\varepsilon, 1-\varepsilon) \} =: \mu_t(\varepsilon)$$

$$\lim_{t \to \infty} \mu_t(\varepsilon) = 0.$$