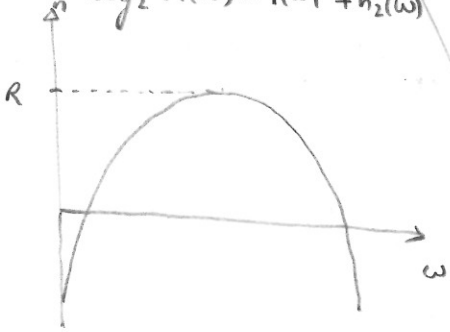
Capacity Via Symmetry I – A New Proof for an Old Code

[M. Mondelli, S. Kudekar, S. Kumar, H.D. Pfister, E. Sasoglu, R. Urbanke]

Basic question of coding theory : how to achieve channel capacity ?

① RANDOM CODES

$\frac{1}{n} \log_2 A(\omega) \sim R - 1 + h_2(\omega)$



create ensemble with pairwise independent codewords and use typicality decoder [Shannon 48]

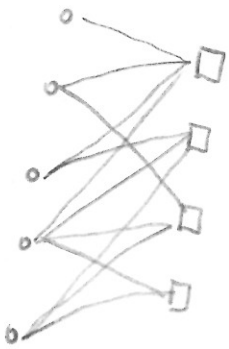
Several classical proofs of this by Feinstein, Elias, Wolfowitz, Gallager.

weight distribution sufficiently close to random one [Poltyrev 94, Shulman-Feder 99]

(SPATIALLY COUPLED)

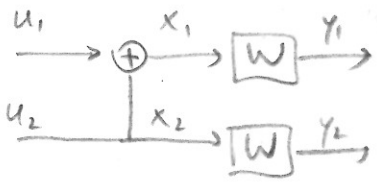
② SPARSE GRAPH CODES : explicitly write down evolution of the decoding process when block length $\to +\infty$ (density evolution)

[Kudekar - Richardson - Urbanke 13]



③ POLAR CODES : concept of channel polarization proof "baked" into the construction of the code

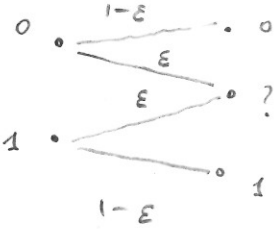
[Arikan 09]



④ <u>SYMMETRY</u> !

Thresholds in Coding Theory

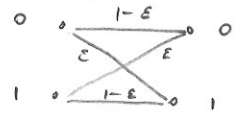
$P_e(n, \varepsilon)$ = error probability for our code of block length n transmitted over a channel with parameter ε.

Think to ε as channel quality (e.g., BEC(ε)) (

)

erasure probability

$P_e(n, \varepsilon)$ increasing in ε.

BSC(ε) 

bit flip probability

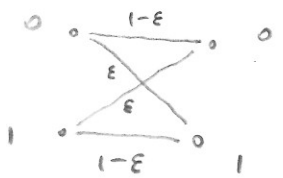
Define $\varepsilon(n, \delta)$ as the channel parameter s.t. $P_e(n, \varepsilon(n,\delta)) = \delta$.

Then, we want that

$$\varepsilon(n, 1-\delta) - \varepsilon(n, \delta) \xrightarrow{n \to +\infty} \emptyset \qquad (\star)$$

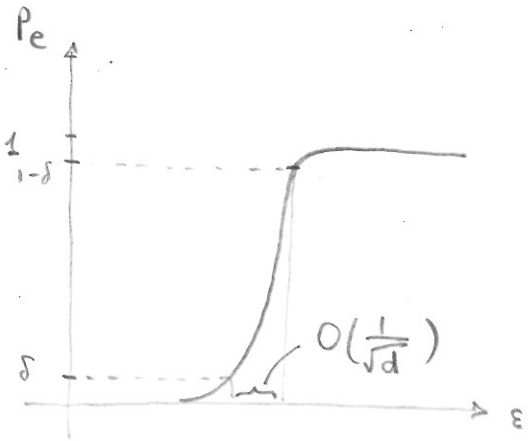
NOTE If we want a code that is capacity-achieving, then the threshold must be at capacity. This follows both from the strong converse of coding theory and from one of our basic techniques.

[Tillich - Zemor 00]: Let $\mathcal{C}$ be a binary, linear code of minimum distance d. Consider transmission over the BEC(ε) or the BSC(ε) and optimal MAP decoding.



Then, $\varepsilon(n, 1-\delta) - \varepsilon(n, \delta) \leq \dfrac{c(\delta)}{\sqrt{d}}$

explicit universal constant depending only on δ.



NOTE This is a VERY strong result!

If the code sequence has linear minimum distance ($d = \alpha n$), then the transition occurs in $O\left(\frac{1}{\sqrt{n}}\right)$.

This is as sharp as it can be since the channel variations are already of order $\frac{1}{\sqrt{n}}$ (e.g. the typical number of erasures/errors is $n\varepsilon \pm c\sqrt{n}$).

EXIT Function and Area Theorem

- Introduced as a visual tool to understand iterative decoding for the BEC [Ashikhmin-Kramer-Ten Brink 04]
- Connection to entropy for general channels (GEXIT function) [Measson-Montanari Richardson-Urbanke 08]

$$\text{EXIT FUNCTION} \quad f(\varepsilon) = \frac{1}{n} \frac{d\, H(\underbrace{\underline{X}}_{(X_1,\dots,X_n)} \mid \underbrace{\underline{Y}}_{(Y_1,\dots,Y_n)})}{d\varepsilon}$$

$$\int_0^1 f(\varepsilon) = \frac{1}{n}\Big[\overbrace{H(\underline{X} \mid \underline{Y}(\varepsilon=1))}^{H(\underline{X}) = nR} - \overbrace{H(\underline{X} \mid \underline{Y}(\varepsilon=0))}^{= 0}\Big] = R$$

The area under the EXIT function is a preserved quantity and does not depend on the code.

Claim
$$f(\varepsilon) = \frac{1}{n}\sum_{i=1}^{n} \overbrace{\mathbb{P}\big(\underbrace{\hat{X}_i^{MAP}(\underbrace{\underline{Y}_{\sim i}}_{(Y_1,\dots,Y_{i-1},Y_{i+1},\dots,Y_n)})}_{\text{MAP estimator of } i\text{-th code bit given observation } \underline{Y}_{\sim i}} = ?\big)}^{f_i(\varepsilon)\text{ - EXIT function associated with } i\text{-th bit}}$$

<u>Proof</u> [Perhaps to be omitted if there is no time]

$$n f(\varepsilon) = \frac{d H(\underline{X} \mid Y_1(\varepsilon), \dots, Y_n(\varepsilon))}{d\varepsilon} = \sum_{i=1}^{n} \frac{\partial H(\underline{X} \mid Y_1(\varepsilon_1), \dots, Y_n(\varepsilon_n))}{\partial \varepsilon_i} \Bigg|_{\substack{\varepsilon_J = \varepsilon \\ \forall J \in [n] \\ = \{1, \dots, n\}}}$$

does not depend on ε_i

$$\overset{=}{\uparrow} \sum_{i=1}^{n} \frac{\partial \left(H(X_i \mid \underline{Y}) + \cancel{H(\underline{X}_{\sim i} \mid X_i, \underline{Y})} \right)}{\partial \varepsilon_i} \Bigg|_{\substack{\varepsilon_J = \varepsilon \\ \forall J \in [n]}}$$

chain rule

$$= \sum_{i=1}^{n} \frac{\partial H(X_i \mid \underline{Y}_{\sim i}, Y_i)}{\partial \varepsilon_i} \Bigg|_{\varepsilon_J = \varepsilon \quad \forall J \in [n]}$$

$$= \sum_{i=1}^{n} \frac{\partial \left[(1-\varepsilon_i) \overset{=\emptyset}{H(X_i \mid Y_i = X_i, \underline{Y}_{\sim i})} + \varepsilon_i \overbrace{H(X_i \mid \underline{Y}_{\sim i}, Y_i = ?)}^{\mathbb{P}(\hat{X}_i^{MAP}(\underline{Y}_{\sim i}) = ?)} \right]}{\partial \varepsilon_i} \Bigg|_{\substack{\varepsilon_J = \varepsilon \\ \forall J \in [n]}}$$

$$= \sum_{i=1}^{n} \mathbb{P}(\hat{X}_i^{MAP}(Y_{\sim i}) = ?) = \sum_{i=1}^{n} f_i(\varepsilon)$$

$\boxtimes$

The Proof!

Ingredients:
1. Symmetry – 2-transitivity
2. Monotone symmetric sets have sharp thresholds
3. EXIT function & Area theorem

(1) $\mathcal{C}$

$$\left.\begin{array}{ccc} \overbrace{x_{11}, \quad \cdots, \quad x_{1n}}^{n \text{ bits}} \\ x_{21}, \quad \cdots, \quad x_{2n} \\ \vdots \qquad\qquad \vdots \\ x_{M1}, \quad \cdots, \quad x_{Mn} \end{array}\right\} M \text{ codewords}$$

bit positions

$\forall \; a, b, c, d \in [n]$ s.t. $a \neq c$, $b \neq d$,

$\exists \; \pi : [n] \to [n]$ so that $\pi(a) = c$, $\pi(b) = d$, $\pi(\mathcal{C}) = \mathcal{C}$

permutation

"For any $a \neq c$, $b \neq d$, there exists a permutation mapping a to c and b to d that leaves the code invariant".

the permutation leaves the code invariant (the permutation of any codeword is still a codeword).

<u>Claim</u> [Kasami – Lin – Peterson 68']

RM codes are 2-transitive.

$\Omega \subseteq \{0,1\}^N$

Ω <u>MONOTONE</u> IF $\omega \geq \omega' \implies \mathbb{1}_\Omega(\omega) \geq \mathbb{1}_\Omega(\omega')$

$\omega_i \geq \omega_i' \quad \forall i \in [N]$

" By adding more 1s, we remain in Ω."

Ω <u>SYMMETRIC</u> IF Ω is 1-transitive

" For any a, b, there exists a permutation mapping a in b that leaves Ω invariant".

This ensures that no single variable has too much influence.

Hamming weight

$$\mu_\varepsilon(\Omega) = \sum_{\omega \in \Omega} \mu_\varepsilon(\omega) = \sum_{\omega \in \Omega} \varepsilon^{W_H(\omega)} (1-\varepsilon)^{N - W_H(\omega)}$$

Bernoulli product measure with parameter ε.

" $\mu_\varepsilon(\Omega)$ is the probability that an i.i.d. $\sim \mathrm{Bern}(\varepsilon)$ vector is in Ω"

[Friedgut - Kalai 96] Claim Let $\Omega \subseteq \{0,1\}^N$ be monotone and symmetric and pick $\delta > 0$. Then, $\mu_\varepsilon(\Omega)$ goes from δ to $1-\delta$ in a window of size $\dfrac{\log(1/\delta)}{\log N}$

Transmission over BEC(ε).

"no erasure"

$\Omega_i \subseteq \{0,1\}^{n-1}$ = set of "bad" erasure patterns for bit i

"erasure". given $Y_{\sim i}$, it is not possible to recover x_i.

EXAMPLE: we receive $0 \; ? \; 1 \; ? \; 0 \; ? \; ?$

Pick $i = 2$, then $\omega = 0\,0\,1\,0\,1\,1$

•) By definition $f_i(\varepsilon) = \mathbb{P}(\hat{X}_i^{MAP}(Y_{\sim i}) = ?) = \mu_\varepsilon(\Omega_i)$

•) Ω_i MONOTONE "more erasures can only hurt"

Proof

ω has more erasures than ω'

$\Updownarrow$

$\omega \succcurlyeq \omega'$

$\Downarrow$

if I cannot decode under ω', I will certain not be able to decode under ω

$\Updownarrow$

$$\mathbb{1}_\Omega(\omega) \geq \mathbb{1}_\Omega(\omega').$$

$\boxtimes$

•) Ω_i SYMMETRIC follows from 2-transitivity of the code

Proof [Perhaps to be omitted if there is no time]

Fix a, b. Then $\exists\, \pi$ (permutation on $[n-1]$) s.t. $\pi(a) = b$ and $\pi(\omega) \in \Omega_i \;\; \forall \omega \in \Omega_i$

this is the claim.

Code is 2-transitive $\Rightarrow \exists\, \pi'$ (permutation on $[n]$) s.t. $\pi'(a) = b$

$\pi'(i) = i$

$\pi'(\mathcal{E}) = \mathcal{E}$

Idea: construct π from π' by removing position i.

$\omega \in \Omega_i \Rightarrow$ there are two compatible codewords c_1, c_2 differing in position i under erasure pattern $\omega \Rightarrow \pi'(c_1), \pi'(c_2)$ are two compatible codewords differing in position i under erasure pattern $\pi(\omega) \Rightarrow \pi(\omega) \in \Omega_i$

$\boxtimes$

INDEPENDENCE

•) $\mu_\varepsilon(\Omega_i) = \mu_\varepsilon(\Omega_J) \quad \forall i, J \in [n]$. follows from 1- transitivity of the code

Proof [Perhaps to be omitted if there is no time]

$\exists\ \bar{u}$ s.t. $\bar{u}(i) = J$ and $\bar{u}(\mathcal{C}) = \mathcal{C}$

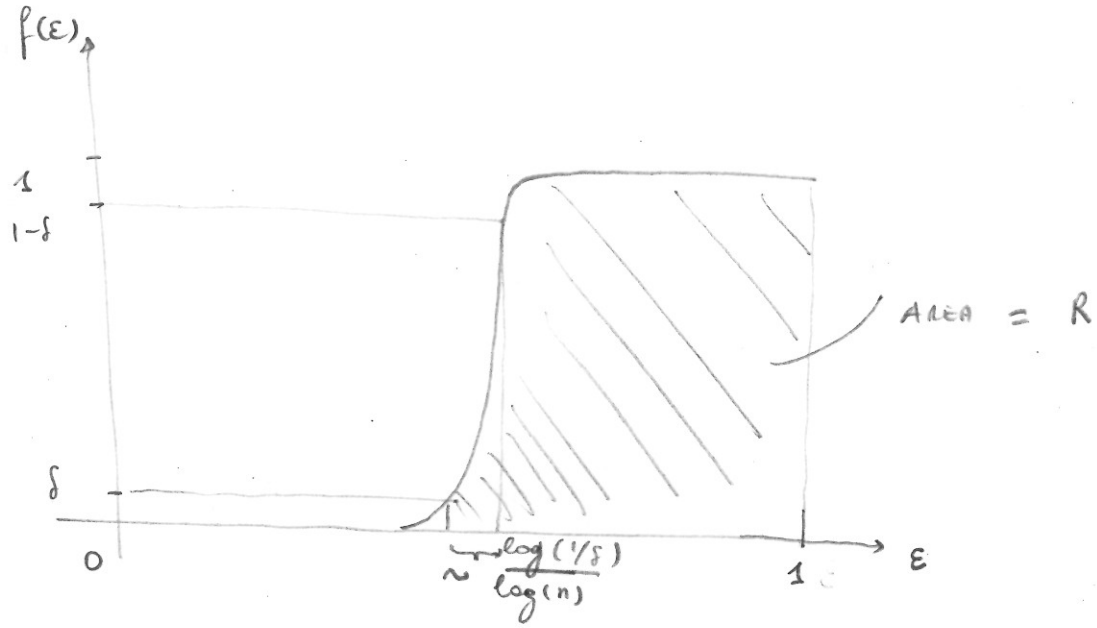
$\omega \in \Omega_i \Rightarrow$ there are two compatible codewords c_1, c_2 differing in position i under erasure pattern $\omega \Rightarrow \bar{u}(c_1), \bar{u}(c_2)$ are two compatible codewords differing in position J under an erasure pattern $\in \Omega_J$ with the same weight as ω.

Thus $\mu_\varepsilon(\Omega_i) \leq \mu_\varepsilon(\Omega_J)$, since different elements of Ω_i are mapped into different elements of Ω_J.

Reverse the role of i and J in the previous argument and then follows ∎

INDEPENDENCE $\Longrightarrow$ $f(\varepsilon) = f_i(\varepsilon) \quad \forall i \in [n]$ (average EXIT function)

Ω_i MONOTONE + SYMMETRIC $\Longrightarrow$ $f(\varepsilon)$ SHARP THRESHOLD



AREA THEOREM $\Longrightarrow$ THRESHOLD IS AT $1-R$

Reliable transmission possible IFF $\varepsilon < 1-R$

$R < 1-\varepsilon = C$

$\Downarrow$

RM codes achieve capacity!