The Unreasonable Effectiveness of Expanders in Coding Theory

Fernando Granha Jeronimo
(UIUC)

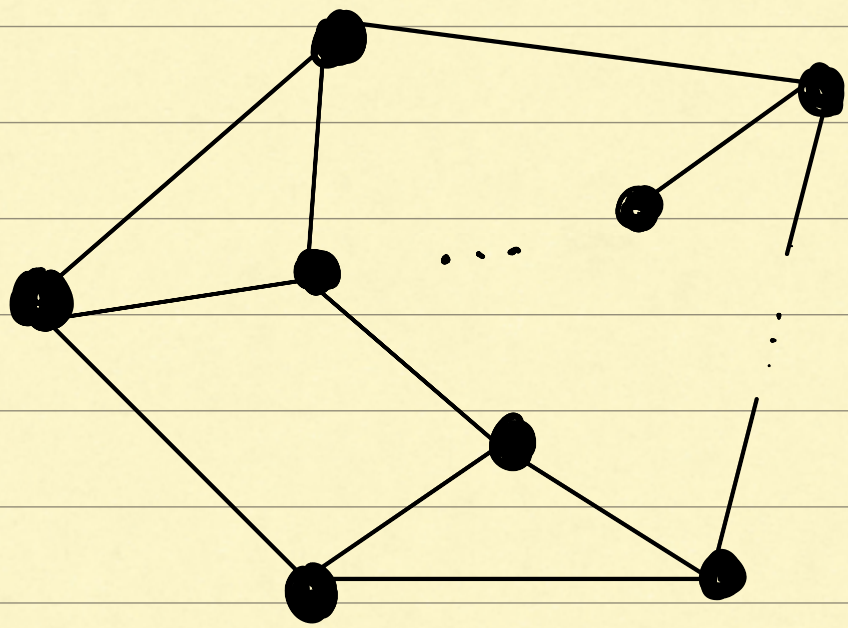
A Glimpse of The Unreasonable Effectiveness of Expanders in Coding Theory

Fernando Granha Jeronimo
(UIUC)

What is one widely studied object in CS/Combinatorics?

What is one widely studied object in CS/Combinatorics?

Graphs $\qquad G = (V, E)$



Expander Graphs

Combine two opposing properties:

Well-connected yet **sparse**

- Mimic the complete graph ✓
- Sparse: can have $O(1)$ degree ✓
- Have (pseudo) random properties ✓
- Can be constructed explicitly ✓

Expander Graphs

Combine two opposing properties:

Well-connected yet sparse

I work with algorithm design.

Why should I care?

Expander Graphs

Combine two opposing properties:

Well-connected yet sparse

I work with algorithm design.

Why should I care?

There is an ongoing revolution
in the design of fast algorithms

Structure-vs-Randomness:

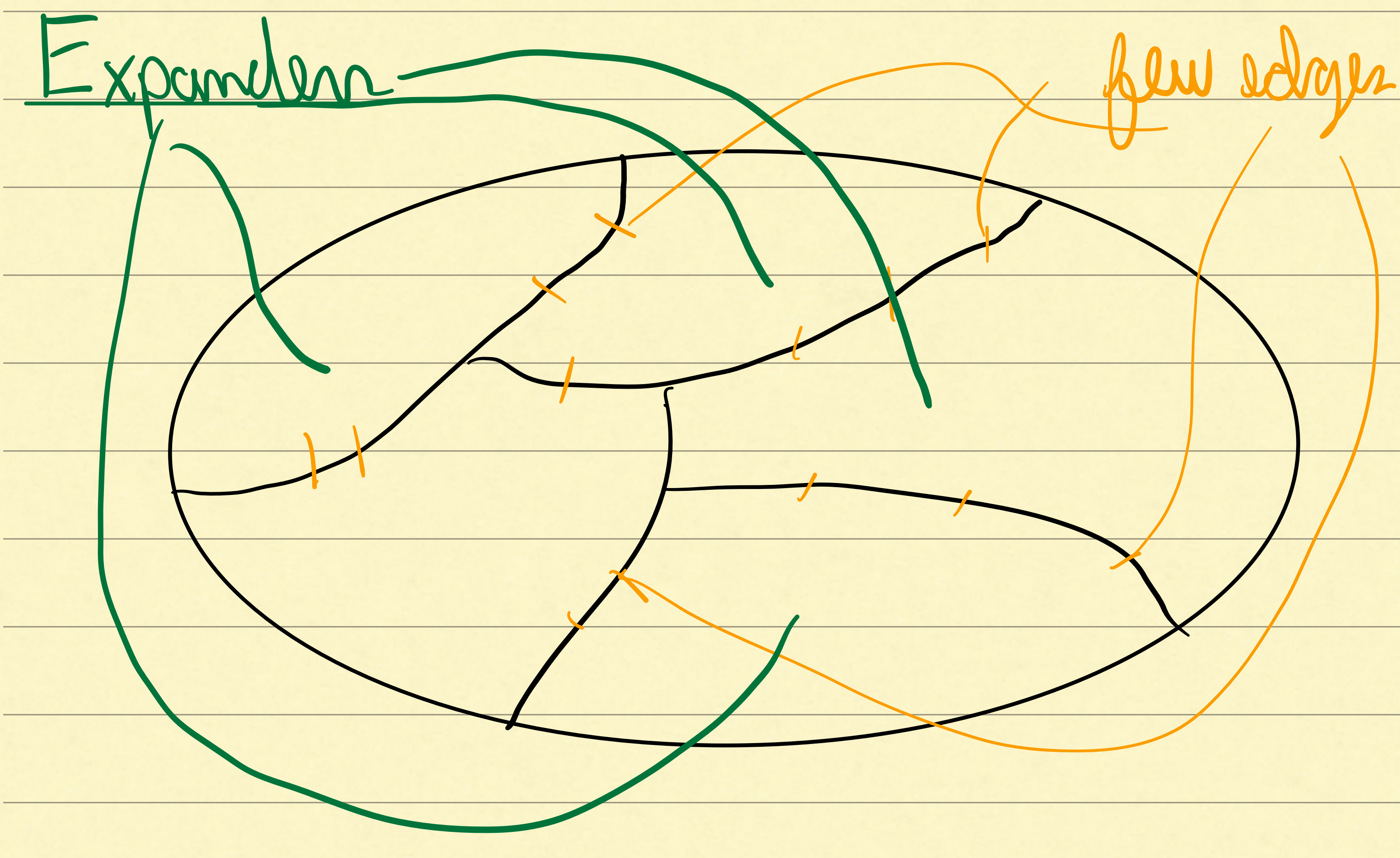
Take an arbitrary $G=(V,E)$ and

decompose it into expanding pieces

Expander Graphs

There is an ongoing revolution in the design of fast algorithms

Structure-vs-Randomness:
Take an arbitrary $G=(V,E)$ and decompose it into expanding pieces



Expander Graphs

Applications to:

- Algorithm Design
- Complexity Theory (PCPs / Hardness)
- Derandomization
- Coding Theory
- MCMC (Mixing / Sampling / Counting)
- Group and Number Theory
- Combinatorics

⋮

... Graphs do seem fundamental.

What about the study of strings?

Σ finite alphabet

$$C \subseteq \Sigma^n$$

... Graphs do seem fundamental.

What about the study of strings?

Σ finite alphabet

$$\mathcal{C} \subseteq \Sigma^n$$

code

Coding Theory: The study of strings

... Graphs do seem fundamental.

What about the study of strings?

Σ finite alphabet

$$\mathcal{C} \subseteq \Sigma^n$$

code

Coding Theory: The study of strings

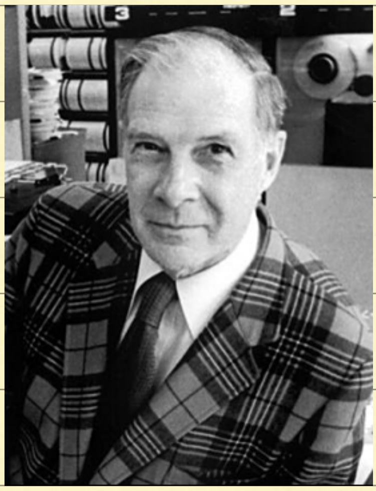
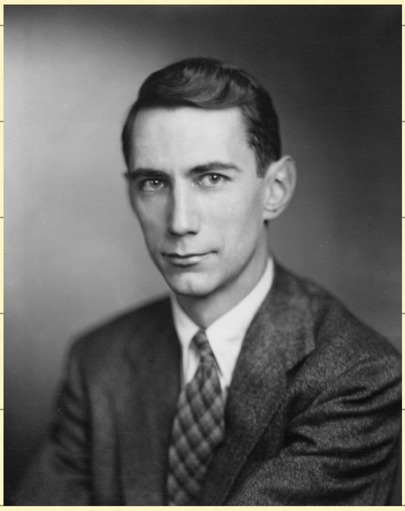
Applications to:

- Communication / Storage
- Complexity Theory / Pseudorandomness
- Quantum Computing / Advantage

⋮

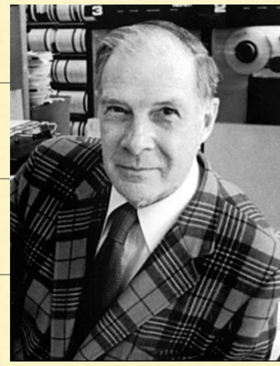
Coding Theory

Two Seminal Models ~1940-1950

Hamming Model	Shannon Model
Errors are adversarial	Errors are Probabilistic
	

Coding Theory

Hamming Model



Errors are adversarial

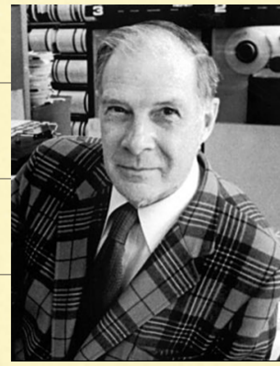
$$\mathcal{C} \subseteq \Sigma^n$$

$$\Delta(\mathcal{C}) = \min_{x,y \in \mathcal{C} : x \neq y} \Delta(x,y) \quad \in [0,1] \quad \text{(distance)}$$

$$r(\mathcal{C}) = \frac{\log_{|\Sigma|}(|\mathcal{C}|)}{n} \quad \in [0,1] \quad \text{(rate)}$$

Coding Theory

Hamming Model



Errors are adversarial

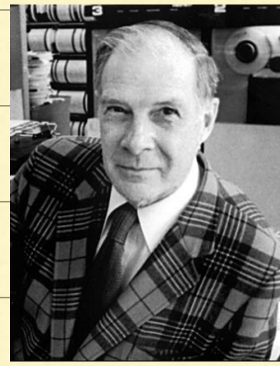
$$\mathcal{C} \subseteq \Sigma^n$$

$$\Delta(\mathcal{C}) = \min_{x,y \in \mathcal{C} : x \neq y} \Delta(x,y) \in [0,1] \quad \text{(distance)}$$

$$r(\mathcal{C}) = \frac{\log_{|\Sigma|}(|\mathcal{C}|)}{n} \in [0,1] \quad \text{(rate)}$$

We want both $\Delta(\mathcal{C})$ and $r(\mathcal{C})$ large, but there is tension

Coding Theory



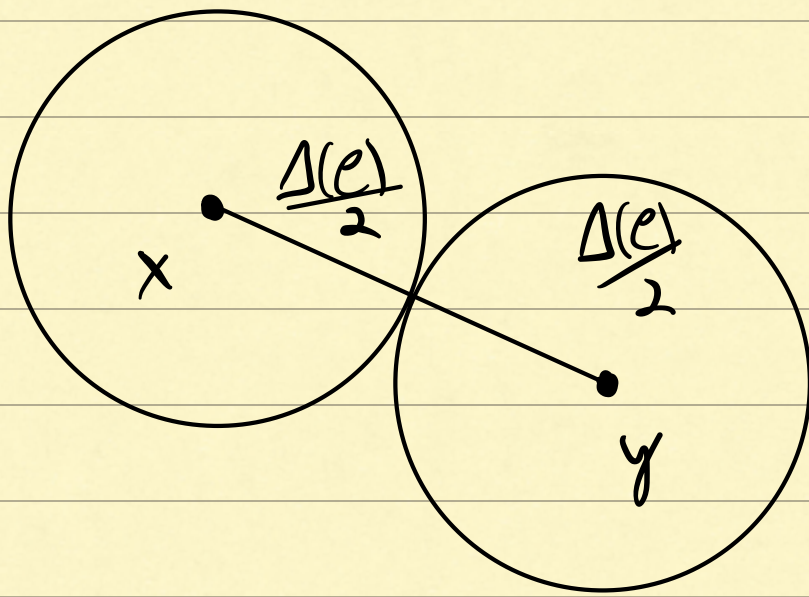
Hamming Model

Errors are adversarial

$$\mathcal{C} \subseteq \Sigma^n$$

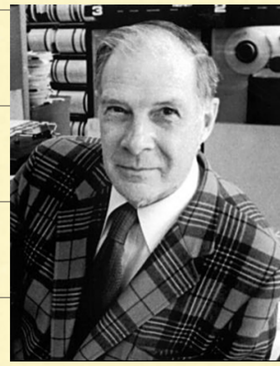
$$\Delta(\mathcal{C}) = \min_{x,y \in \mathcal{C};\, x \neq y} \Delta(x,y) \in [0,1] \quad \text{(distance)}$$

$$r(\mathcal{C}) = \frac{\log_{|\Sigma|}(|\mathcal{C}|)}{n} \in [0,1] \quad \text{(rate)}$$



Coding Theory

Hamming Model



Errors are adversarial

$$\mathcal{C} \subseteq \Sigma^n$$

$$\Delta(\mathcal{C}) = \min_{x,y \in \mathcal{C} : x \neq y} \Delta(x,y) \in [0,1] \quad \text{(distance)}$$

$$r(\mathcal{C}) = \frac{\log_{|\Sigma|}(|\mathcal{C}|)}{n} \in [0,1] \quad \text{(rate)}$$

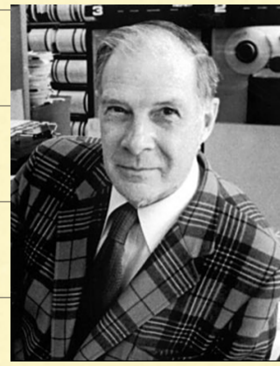
Hamming model

1950 | 1950s

Random codes
Gilbert-Varshamov

Coding Theory

Hamming Model



Errors are adversarial

$$\mathcal{C} \subseteq \Sigma^n$$

$$\Delta(\mathcal{C}) = \min_{x,y \in \mathcal{C};\, x \neq y} \Delta(x,y) \quad \in [0,1] \quad \text{(distance)}$$

$$r(\mathcal{C}) = \frac{\log_{|\Sigma|}(|\mathcal{C}|)}{n} \quad \in [0,1] \quad \text{(rate)}$$

How to construct explicit codes?

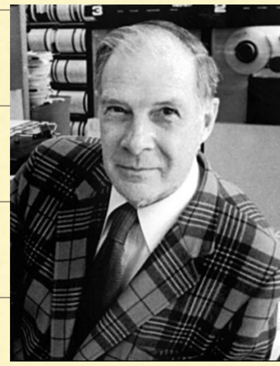
Hamming model

1950

1950s

Random codes
Gilbert-Varshamov

Coding Theory



Hamming Model

Errors are adversarial

$$\mathcal{C} \subseteq \Sigma^n$$

$$\Delta(\mathcal{C}) = \min_{x,y \in \mathcal{C}:\, x \neq y} \Delta(x,y) \in [0,1] \text{ (distance)}$$

$$r(\mathcal{C}) = \frac{\log_{|\Sigma|}(|\mathcal{C}|)}{n} \in [0,1] \text{ (rate)}$$

How to construct <u>explicit</u> codes?

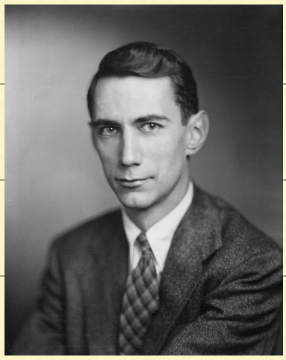
Hamming model — 1950

1950s — Random codes, Gilbert-Varshamov

Ta-Shma's codes — 2017

Coding Theory

Shannon Model



Errors are Probabilistic

Ex binary symmetric channel BSC_p

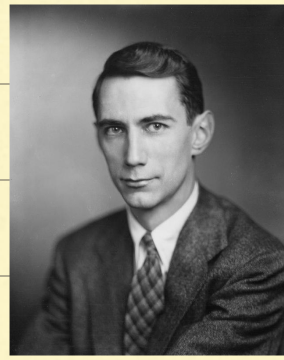
Each bit is independently flipped w.p. $p \in [0, 1]$

Using Probabilistic Method Shannon proved

Thm [Capacity] $\forall\, \varepsilon > 0,\ p < 1 - \frac{1}{q}$

rate $\geq 1 - H_q(p) - \varepsilon$ (but no more)

Coding Theory



Shannon Model

Errors are Probabilistic

Using Probabilistic Method Shannon proved

Thm [Capacity] $\forall\, \varepsilon > 0,\ \rho < 1 - \frac{1}{q}$

$$\text{rate} \geq 1 - H_q(\rho) - \varepsilon \quad \text{(but no more)}$$

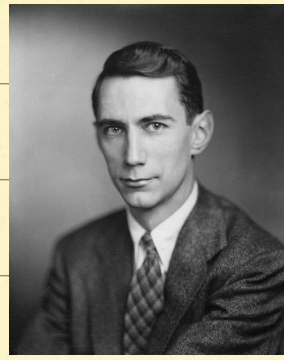
How to construct explicit capacity achieving codes?

Shannon's random codes

1948

Coding Theory

Shannon Model



Errors are Probabilistic

Using Probabilistic Method Shannon proved

Thm [Capacity] $\forall\, \varepsilon > 0$, $\rho < 1 - \frac{1}{q}$

$$\text{rate} \geq 1 - H_q(\rho) - \varepsilon \quad \text{(but no more)}$$

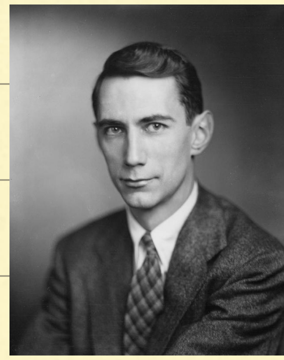
How to construct explicit capacity achieving codes?

Shannon's random codes — 1948

Arikan's polar codes — 2006

Coding Theory

Shannon Model



Errors are Probabilistic

Using Probabilistic Method Shannon proved

Thm [Capacity] $\forall\, \varepsilon > 0$, $p < 1 - \frac{1}{q}$

$$\text{rate} \geq 1 - H_q(p) - \varepsilon \quad \text{(but no more)}$$

How to construct explicit capacity achieving codes?

Shannon's random codes — 1948

Arikan's polar codes — 2006

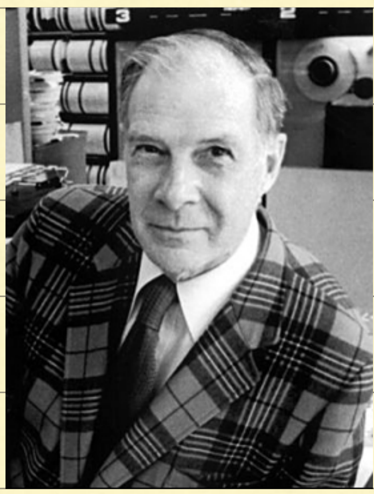
5G

Coding Theory

Two Seminal Models ~1940-1950

Hamming Model

Errors are adversarial

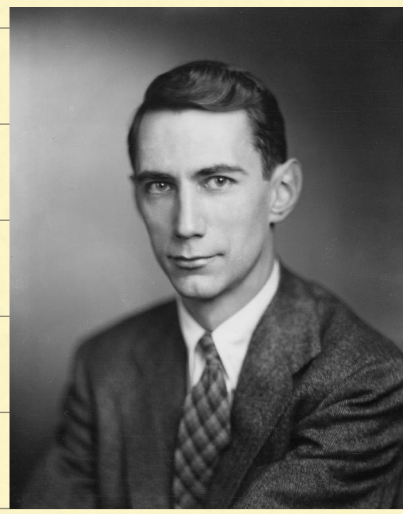


More general errors ✓

Worse Parameters ✗

Shannon Model

Errors are Probabilistic

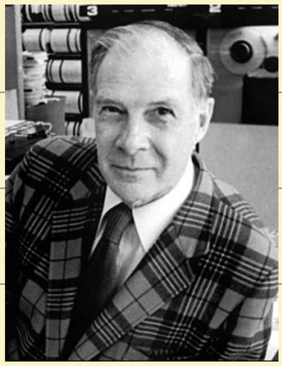


Simpler errors ✗

Better Parameters ✓

Coding Theory

Hamming Model



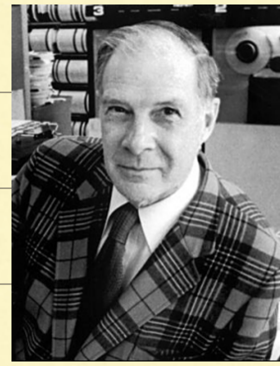
Errors are adversarial

$$\mathcal{E} \subseteq \Sigma^n \qquad \rho = \Delta(\mathcal{E})$$

How can we decode from $> \rho/2$ adversarial errors?

Coding Theory

Hamming Model



Errors are adversarial

$$\mathcal{C} \subseteq \Sigma^n \qquad \rho = \Delta(\mathcal{C})$$

How can we decode from $> \rho/2$ adversarial errors?

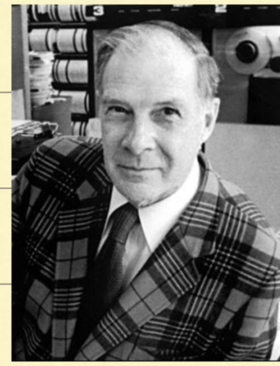
Impossible to unique decode!

Possible if we allow a list of codewords

[Elias'57, Wozencraft'58]

Coding Theory

Hamming Model



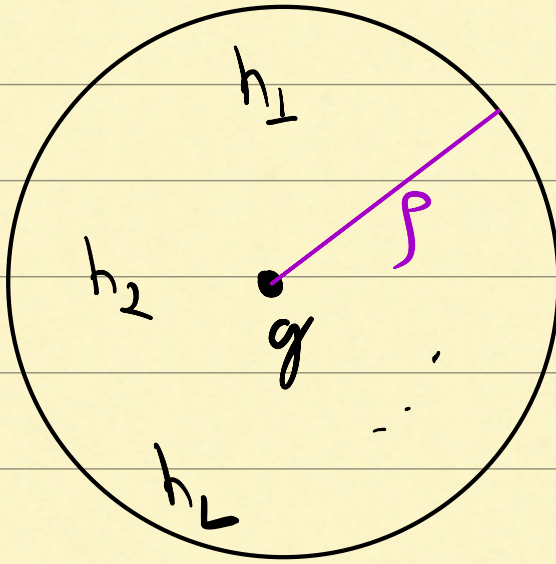
Errors are adversarial

$$\mathcal{C} \subseteq \Sigma^n$$

List Decoding [Elias '57, Wozencraft '58]

$\mathcal{C}$ is (ρ, L)-list decodable if

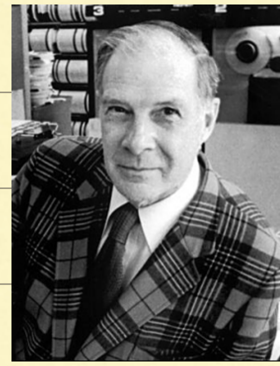
$$\forall y \in \Sigma^n, \ |\mathrm{Ball}(y, \rho) \cap \mathcal{C}| \le L.$$



Coding Theory

Hamming Model

Errors are adversarial



$$\mathcal{C} \subseteq \Sigma^n$$

List Decoding [Elias'57, Wozencraft'58]

$\mathcal{C}$ is (ρ, L)-list decodable if

$$\forall\, y \in \Sigma^n, \ |\mathrm{Ball}(y,\rho) \cap \mathcal{C}| \le L.$$

Zyablov-Pinsker'81 via Probabilistic Method

Thm [List Decoding Capacity]

$$\forall\, \varepsilon > 0, \ \rho \in (0, 1 - 1/q)$$

(1) $\exists\, \mathcal{C} \subseteq \mathbb{F}_q^n$, (ρ, L)-list decodable

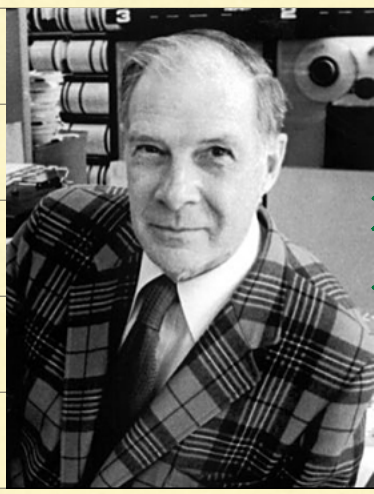
with $r(\mathcal{C}) \geq 1 - H_q(\rho) - \varepsilon$ and $L = O\left(\frac{1}{\varepsilon}\right)$.

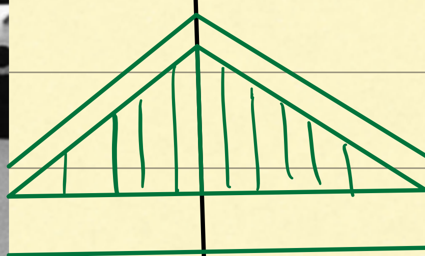
(best possible)

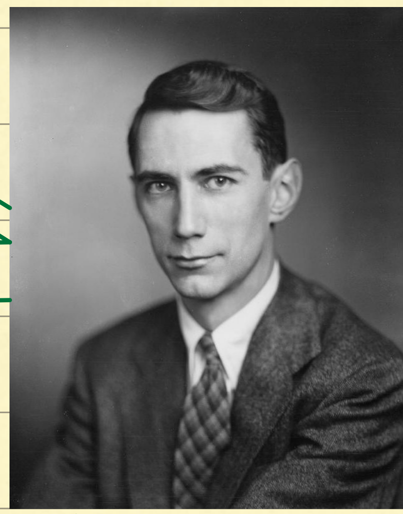
Coding Theory

Two Seminal Models ~1940-1950

Hamming Model	Shannon Model
Errors are adversarial	Errors are Probabilistic







list decoding

More general errors ✓	Simpler errors ^
Worse Parameters ^	Better Parameters ✓

Quest to Construct Explicit Codes achieving List Decoding Capacity

Hamming Model

Errors are adversarial

Shannon Model

Errors are Probabilistic

[Pernice-Sprumont-Wootters '25]

list decoding capacity $\overset{*}{\Longrightarrow}$ capacity

Quest to Construct Explicit Codes achieving List Decoding Capacity

Hamming Model	Shannon Model
Errors are adversarial	Errors are Probabilistic

[Pernice-Sprumont-Wootters '25]

list decoding capacity $\overset{*}{\Longrightarrow}$ capacity

$\not\Longleftarrow$ not always true

Singleton Bound

Thm [1950's]
$\forall$ code $\mathcal{C}$ with $r(\mathcal{C}) = R$, we have
$$\Delta(\mathcal{C}) \leq 1 - R$$

Reed - Solomon (RS) codes achieve this bound

$\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{F}$ distinct

$$p(x) = c_0 + c_1 x + \ldots + c_{k-1} x^{k-1}$$

$$\Big\downarrow \text{ Encoding map}$$

$$(p(\alpha_1), p(\alpha_2), \ldots, p(\alpha_n))$$

codeword

Singleton Bound

Thm

$\forall$ code $\mathcal{C}$ with $r(\mathcal{C}) = R$, we have

$\Delta(\mathcal{C}) \leq 1 - R$

Reed-Solomon (RS) codes achieve this bound

$\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{F}$ distinct

$$p(x) = c_0 + c_1 x + \ldots + c_{k-1} x^{k-1}$$

$$\Big\downarrow \text{ Encoding map}$$

$$(p(\alpha_1), p(\alpha_2), \ldots, p(\alpha_n))$$

codeword

$\boxed{\text{Drawback: } |\mathbb{F}| \geq n \text{ large alphabet}}$

List Decoding with Explicit Codes

Efficient list decoding for RS

$1-\sqrt{2R}$ [Sudan '95]

$1-\sqrt{R}$ [Guruswami-Sudan '97]
(Johnson bound)

Parvaresh-Vardy '05
$1-O(R\log 1/R)$

Folded Reed-Solomon (FRS)

$\boxed{1-R-\varepsilon}$ [Guruswami-Rudra '06]
(List decoding capacity)

Amazing Synergy

Expanders + Codes

Recent breakthroughs:

Ta-Shma's Codes '17

[Explicit Binary Codes near GV]

[+60 years after random construction]

c^3-LTC

[Dinur et al. '22]

Good Quantum LDPC Codes

[Panteleev - Kalachev '22]

Warm-up: Code Concatenation

$\mathcal{C}_{out} \subseteq \Sigma_{out}^{n}$

$\mathcal{C}_{in} \subseteq \Sigma_{in}^{d}$ with $|\mathcal{C}_{in}| = |\Sigma_{out}|$

$\mathcal{C}_{in} \simeq \Sigma_{out}$

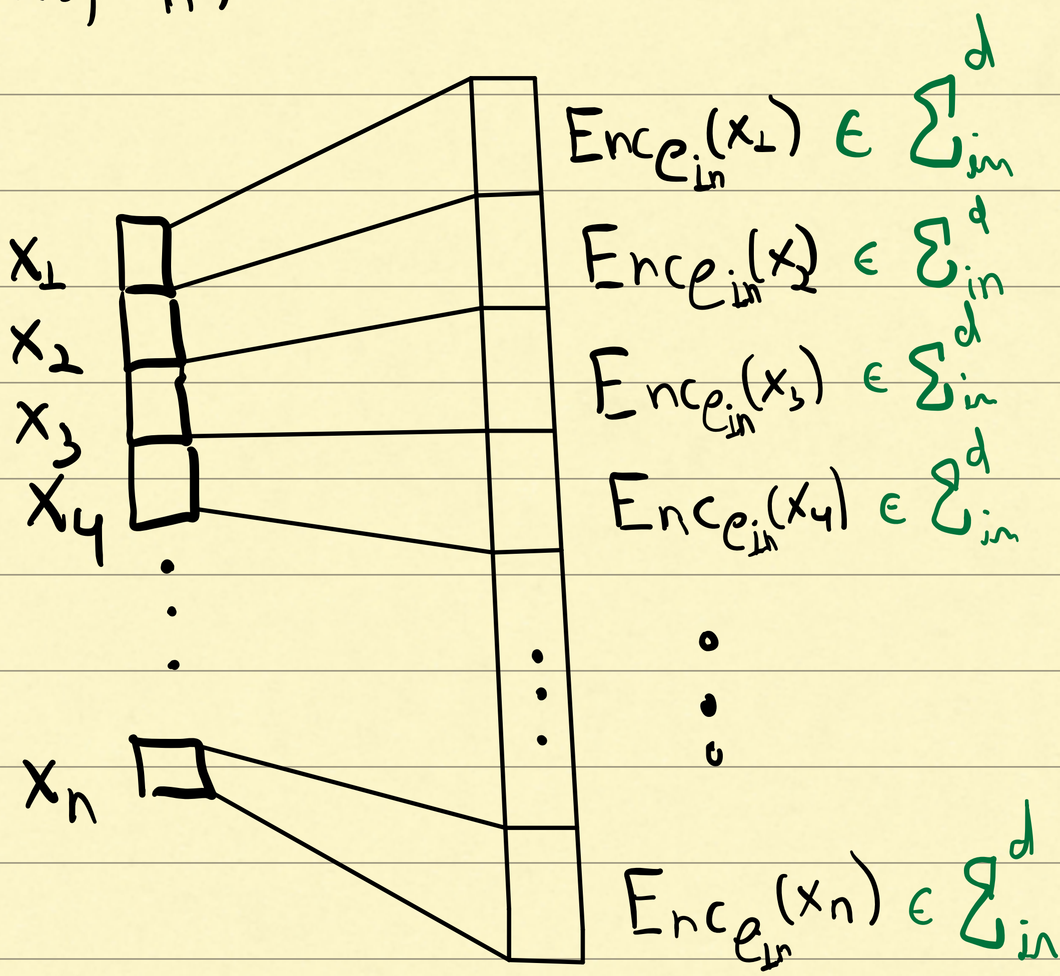
Code Concatenation

$$x \in \mathcal{C}_{out} \subseteq \Sigma_{out}^{n}$$

$$\mathcal{C}_{in} \subseteq \Sigma_{in}^{d} \quad \text{with} \quad |\mathcal{C}_{in}| = |\Sigma_{out}|$$

$$\mathcal{C}_{in} \simeq \Sigma_{out}$$

$$x = (x_1, x_2, \dots, x_n)$$



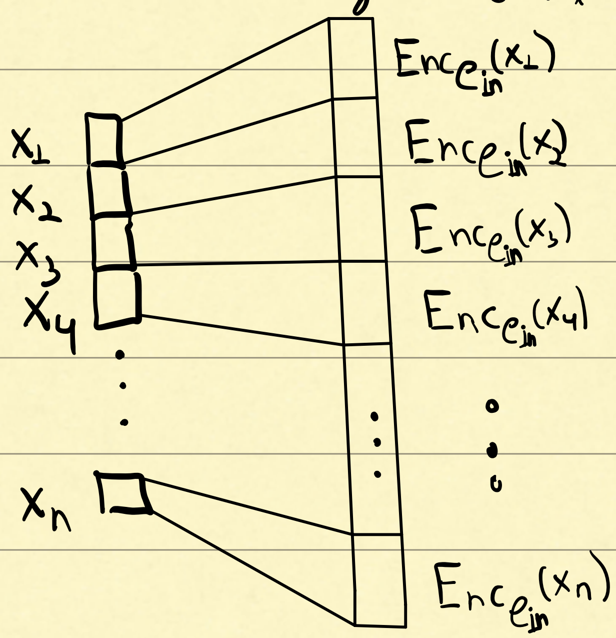
Code Concatenation

$$x \in \mathcal{C}_{out} \subseteq \Sigma_{out}^{n}$$

$$\mathcal{C}_{in} \subseteq \Sigma_{in}^{d} \quad \text{with} \quad |\mathcal{C}_{in}| = |\Sigma_{out}|$$

$x = (x_1, x_2, \dots, x_n)$

$y \in (\Sigma_{in})^{nd}$



$$\delta(\mathcal{C}_{out} \circ \mathcal{C}_{in}) = \delta(\mathcal{C}_{out})\,\delta(\mathcal{C}_{in})$$

$$r(\mathcal{C}_{out} \circ \mathcal{C}_{in}) = r(\mathcal{C}_{out})\,r(\mathcal{C}_{in})$$

Code Concatenation

$$\delta(\mathcal{C}_{out} \circ \mathcal{C}_{in}) = \delta(\mathcal{C}_{out})\,\delta(\mathcal{C}_{in})$$

$$r(\mathcal{C}_{out} \circ \mathcal{C}_{in}) = r(\mathcal{C}_{out})\,r(\mathcal{C}_{in})$$

Dream Goal: Another code operation $\square$

$$\delta(\mathcal{C}_{out} \,\square\, \mathcal{C}_{in}) \simeq \delta(\mathcal{C}_{in})$$

$$r(\mathcal{C}_{out} \,\square\, \mathcal{C}_{in}) \simeq r(\mathcal{C}_{in})$$

The "good" properties of a small <u>local code</u> $\mathcal{C}_{in}$ become the properties of the <u>global code</u> $(\mathcal{C}_{out} \square \mathcal{C}_{in})$

Local-to-Global

Dream Goal: Another code operation $\square$

$$\delta(\mathcal{C}_{out} \square \mathcal{C}_{in}) \simeq \delta(\mathcal{C}_{in})$$

$$r(\mathcal{C}_{out} \square \mathcal{C}_{in}) \simeq r(\mathcal{C}_{in})$$

The "good" properties of a small local code $\mathcal{C}_{in}$ become the properties of the global code $(\mathcal{C}_{out} \square \mathcal{C}_{in})$

[Alon-Edmonds-Luby '94]

This is possible using expander graphs!

AEL Construction

AEL Construction

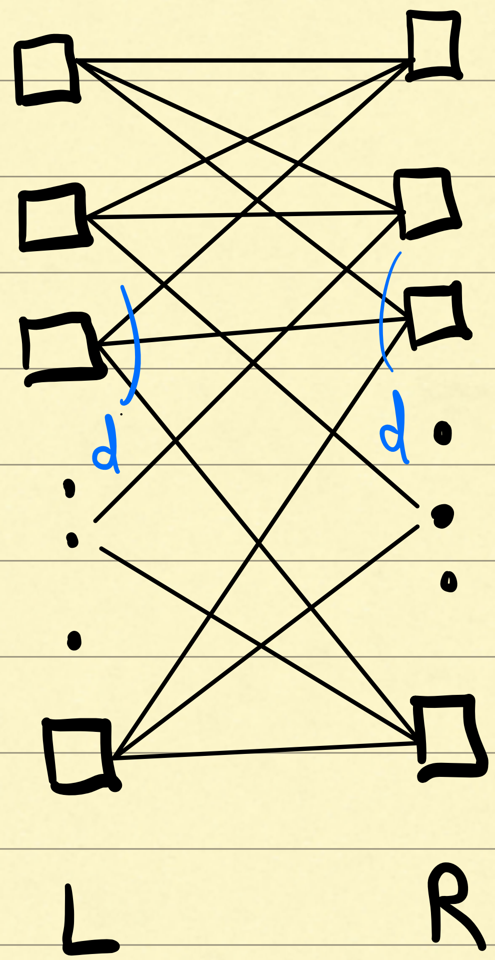
$AEL(G, C_{out}, C_{in})$

$G = (L, R, E)$ $|L| = |R| = n$ d-regular

AEL Construction

$AEL(G, \mathcal{C}_{out}, \mathcal{C}_{in})$

$G = (L, R, E)$ bipartite $|L| = |R| = n$ d-regular



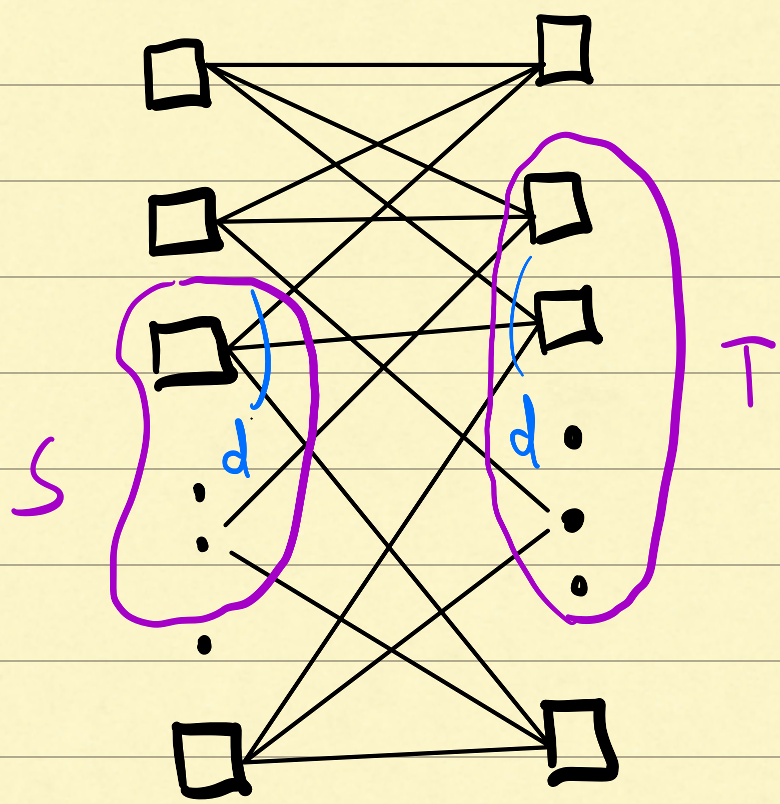
AEL Construction

AEL$(G, \mathcal{C}_{out}, \mathcal{C}_{in})$

$G = (L, R, E)$ bipartite $|L| = |R| = n$ d-regular

λ-expander $\lambda \in [0, 1)$

$$\boxed{\left| E(S,T) - \frac{d}{n}|S||T| \right| \le \lambda d n}$$



$\forall$ $S \subseteq L$ $T \subseteq R$

AEL Construction

$\mathrm{AEL}(G, \mathcal{C}_{out}, \mathcal{C}_{in})$

$G = (L, R, E)$ bipartite $|L| = |R| = n$ d-regular

$\mathcal{C}_{out} \subseteq \Sigma_{out}^{n}$, $\mathcal{C}_{in} \subseteq \Sigma_{in}^{d}$ with $|\Sigma_{out}| = |\mathcal{C}_{in}|$

$\Sigma_{out} \simeq \mathcal{C}_{in}$

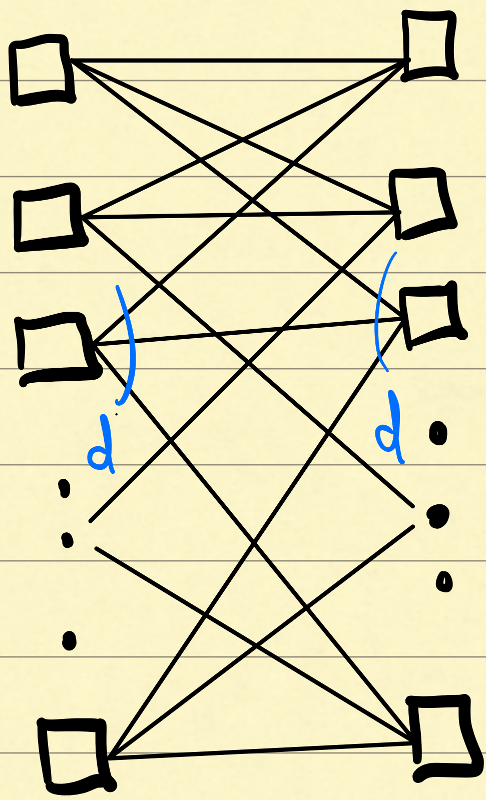
$x = (x_1, \ldots, x_n) \in \mathcal{C}_{out}$

$x_1 \to \mathrm{Enc}_{\mathcal{C}_{in}}(x_1)$

$x_2 \to \mathrm{Enc}_{\mathcal{C}_{in}}(x_2)$

$x_3 \to \mathrm{Enc}_{\mathcal{C}_{in}}(x_3)$

$x_n \to \mathrm{Enc}_{\mathcal{C}_{in}}(x_n)$



L R

AEL Construction

$\mathrm{AEL}(G, \mathcal{C}_{out}, \mathcal{C}_{in})$

$G = (L, R, E)$ bipartite $|L| = |R| = n$ d-regular

$\mathcal{C}_{out} \subseteq \Sigma_{out}^{n}$, $\mathcal{C}_{in} \subseteq \Sigma_{in}^{d}$ with $|\Sigma_{out}| = |\mathcal{C}_{in}|$

$\Sigma_{out} \simeq \mathcal{C}_{in}$

$x = (x_1, \dots, x_n) \in \mathcal{C}_{out}$

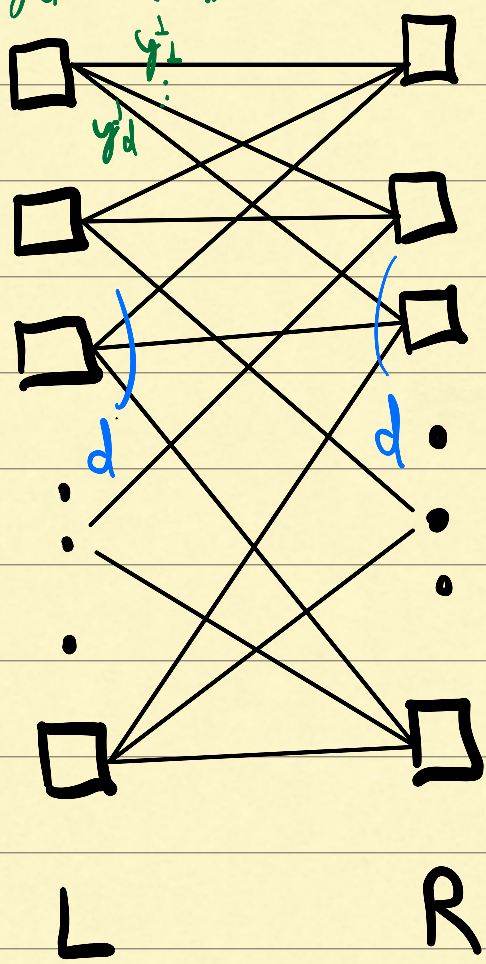
$(y^1_1, y^1_2, \dots, y^1_d) \in (\Sigma_{in})^d$

$\shortparallel$

$x_1 \to \mathrm{Enc}_{\mathcal{C}_{in}}(x_1)$

$x_2 \to \mathrm{Enc}_{\mathcal{C}_{in}}(x_2)$

$x_3 \to \mathrm{Enc}_{\mathcal{C}_{in}}(x_3)$



$x_n \to \mathrm{Enc}_{\mathcal{C}_{in}}(x_n)$

$\in (\Sigma_{in})^d$

AEL Construction

$AEL(G, \mathcal{C}_{out}, \mathcal{C}_{in})$

$G = (L, R, E)$ bipartite $|L| = |R| = n$ d-regular

$\mathcal{C}_{out} \subseteq \Sigma_{out}^{n}$, $\mathcal{C}_{in} \subseteq \Sigma_{in}^{d}$ with $|\Sigma_{out}| = |\mathcal{C}_{in}|$

$\Sigma_{out} \simeq \mathcal{C}_{in}$

$x = (x_1, \dots, x_n) \in \mathcal{C}_{out}$

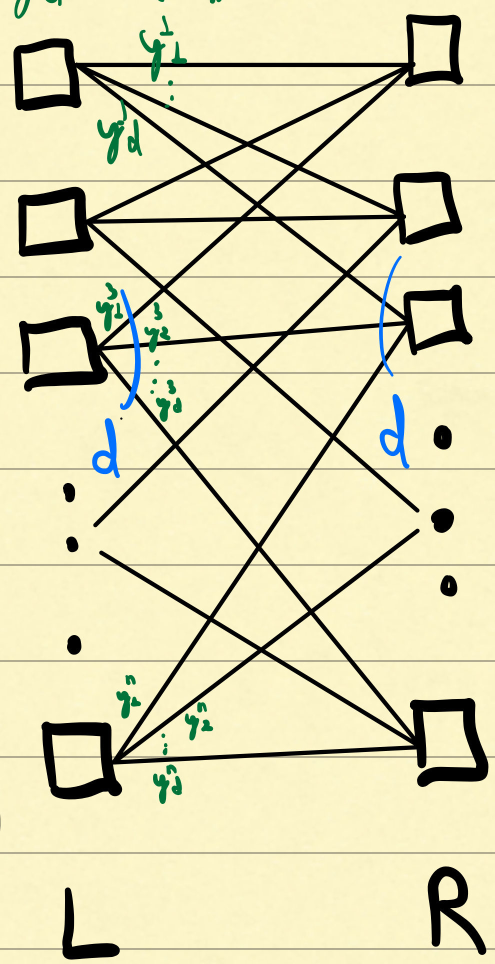
$(y_1^1, y_2^1, \dots, y_d^1) \in (\Sigma_{in})^d$

$\|$

$x_1 \rightarrow Enc_{\mathcal{C}_{in}}(x_1)$

$x_2 \rightarrow Enc_{\mathcal{C}_{in}}(x_2)$

$x_3 \rightarrow Enc_{\mathcal{C}_{in}}(x_3)$



$x_n \rightarrow Enc_{\mathcal{C}_{in}}(x_n)$

$\cap$

$(\Sigma_{in})^d$

AEL Construction

$\text{AEL}(G, \mathcal{C}_{out}, \mathcal{C}_{in})$

$G = (L, R, E)$ bipartite $|L| = |R| = n$ d-regular

$\mathcal{C}_{out} \subseteq \Sigma_{out}^{n}$, $\mathcal{C}_{in} \subseteq \Sigma_{in}^{d}$ with $|\Sigma_{out}| = |\mathcal{C}_{in}|$

$\Sigma_{out} \simeq \mathcal{C}_{in}$

$x = (x_1, \dots, x_n) \in \mathcal{C}_{out}$

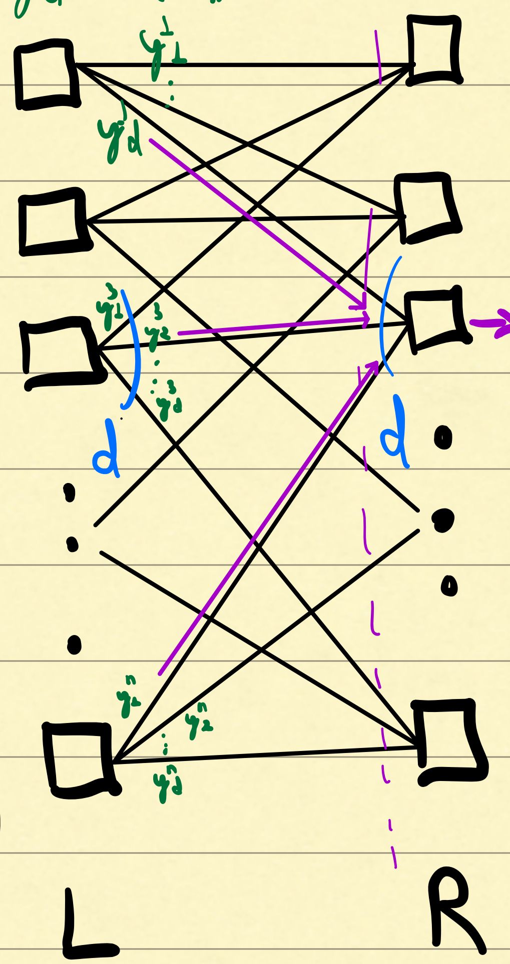
$(y_1^1, y_2^1, \dots, y_d^1) \in (\Sigma_{in})^d$

$\shortparallel$

$x_1 \to \text{Enc}_{\mathcal{C}_{in}}(x_1)$

$x_2 \to \text{Enc}_{\mathcal{C}_{in}}(x_2)$

$x_3 \to \text{Enc}_{\mathcal{C}_{in}}(x_3)$



Collect Symbols

$(y_d^1, y_2^3, \dots, y_d^n)$

$\cap$

$(\Sigma_{in})^d$

$x_n \to \text{Enc}_{\mathcal{C}_{in}}(x_n)$

$\cap$

$(\Sigma_{in})^d$

AEL Construction

$\mathrm{AEL}(G, \mathcal{C}_{out}, \mathcal{C}_{in})$

$G = (L, R, E)$ bipartite $|L| = |R| = n$ d-regular

$\mathcal{C}_{out} \subseteq \Sigma_{out}^{n}$, $\mathcal{C}_{in} \subseteq \Sigma_{in}^{d}$ with $|\Sigma_{out}| = |\mathcal{C}_{in}|$

$\Sigma_{out} \simeq \mathcal{C}_{in}$

$x = (x_1, \ldots, x_n) \in \mathcal{C}_{out}$

$(y^1_1, y^1_2, \ldots, y^1_d) \in (\Sigma_{in})^d$

$\quad\|$

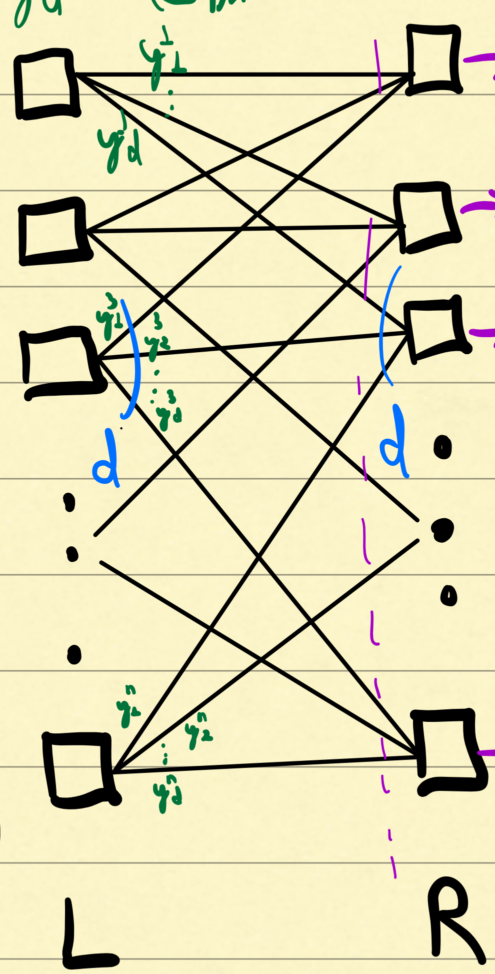
$x_1 \to \mathrm{Enc}_{\mathcal{C}_{in}}(x_1)$

$x_2 \to \mathrm{Enc}_{\mathcal{C}_{in}}(x_2)$

$x_3 \to \mathrm{Enc}_{\mathcal{C}_{in}}(x_3)$

$x_n \to \mathrm{Enc}_{\mathcal{C}_{in}}(x_n)$

$\in (\Sigma_{in})^d$



Collect Symbols

$\to y_1 \in (\Sigma_{in})^d$

$\to y_2$

$\to y_3 = (y^1_d, y^3_2, \ldots, y^n_d)$

$\in (\Sigma_{in})^d$

$\to y_n$ Final codeword in $(\Sigma_{in}^d)^R$

How can we analyze the distance?

$$g, g' \in \mathcal{C}_{AEL} \subseteq \left(\Sigma_{in}^{d}\right)^{R}$$

$$g \neq g'$$

How can we analyze the distance?

$$g, g' \in \mathcal{C}_{AEL} \subseteq (\Sigma_{in}^{d})^{R}$$

$$g \neq g'$$

$$x, x' \in \mathcal{C}_{out} \qquad \Delta(x, x') \geq \delta(\mathcal{C}_{out})$$

$i \in S$ iff $x_i \neq x'_i$

$S \subseteq L$ R

$T = N(S) := \{t \in R \mid \exists \underset{s}{(\overset{\in}{s}, t)} \in E$ witness a disagreement$\}$

How can we analyze the distance?

$$g, g' \in \mathcal{C}_{AEL} \subseteq (\Sigma_{in}^{d})^{R}$$

$$g \neq g'$$

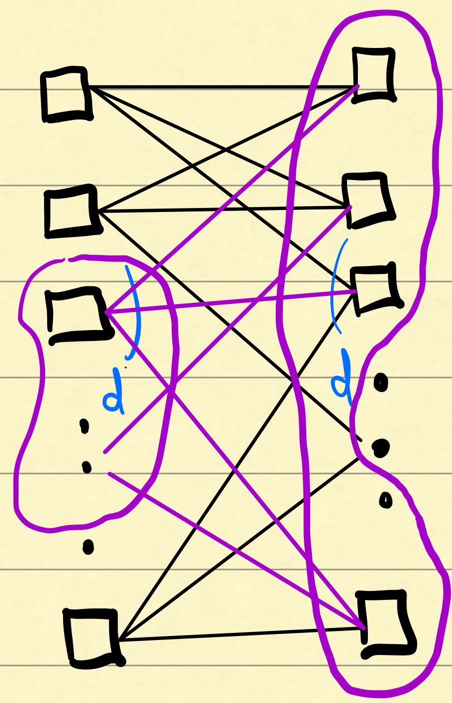
$$x, x' \in \mathcal{C}_{out} \qquad \Delta(x, x') \geq \delta_{out}$$

$$\Delta(g, g') \geq |T|/n$$

$i \in S$ iff $x_i \neq x'_i$

$|S| \geq \delta_{out} \cdot n$

S



$T = N(S)$

$S \subseteq L$ $\quad$ R

How can we analyze the distance?

$g, g' \in \mathcal{C}_{AEL} \subseteq (\Sigma_{in}^{d})^{R}$

$g \neq g'$

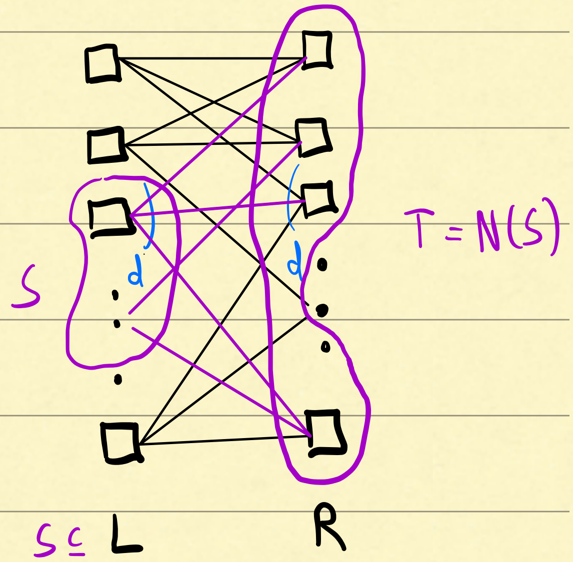
$x, x' \in \mathcal{C}_{out}$

$\Delta(x, x') \geq \delta_{out}$

$\Delta(g, g') \geq |T|/n$

$i \in S$ iff $x_i \neq x'_i$

$|S| \geq \delta_{out} \cdot n$



$$\boxed{\left| E(S,T) - \frac{d}{n}|S||T| \right| \leq \lambda d n}$$

$$E(S,T) \leq \frac{d}{n}|S||T| + \lambda d n$$

How can we analyze the distance?

$g, g' \in \mathcal{C}_{AEL} \subseteq (\Sigma_{in}^{d})^{R}$

$g \neq g'$

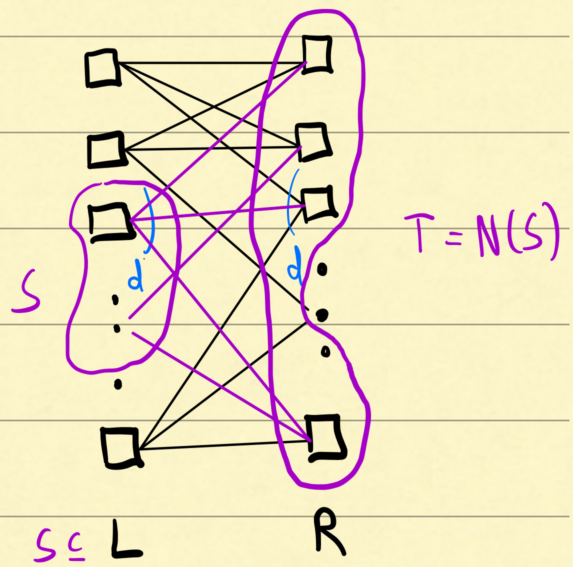
$x, x' \in \mathcal{C}_{out}$

$\Delta(x, x') \geq \delta_{out}$

$\Delta(g, g') \geq |T|/n$

$i \in S$ iff $x_i \neq x_i'$

$|S| \geq \delta_{out} \cdot n$



$$\boxed{\left| E(S,T) - \frac{d}{n}|S||T| \right| \leq \lambda d n}$$

$$\delta_{in} d |S| \leq E(S,T) \leq \frac{d}{n} |S||T| + \lambda d n$$

How can we analyze the distance?

$g, g' \in \mathcal{C}_{AEL} \subseteq (\Sigma_{in}^{d})^{R}$

$g \neq g'$

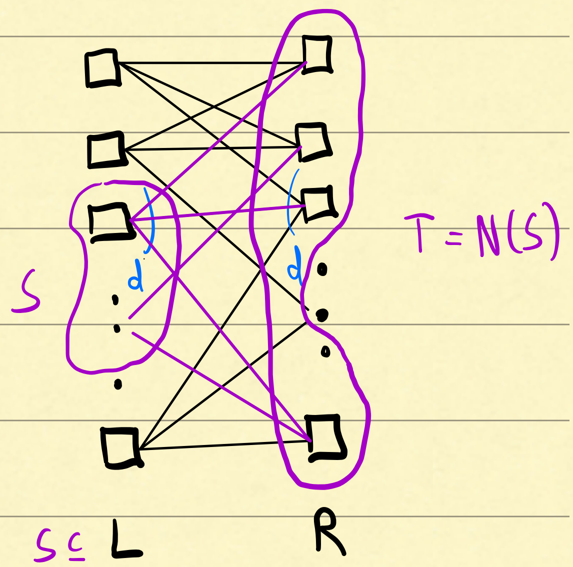
$x, x' \in \mathcal{C}_{out}$

$\Delta(x, x') \geq \delta_{out}$

$\Delta(g, g') \geq |T|/n$

$i \in S$ iff $x_i \neq x'_i$

$|S| \geq \delta_{out} \cdot n$



$$\boxed{\left| E(S,T) - \frac{d}{n} |S||T| \right| \leq \lambda d n}$$

$$\delta_{in} d |S| \leq E(S,T) \leq \frac{d}{n} |S||T| + \lambda d n$$

$$\Downarrow$$

$$\delta_{in} - \frac{\lambda n}{|S|} \leq \frac{|T|}{n} \qquad \boxed{|S| \geq \delta_{out} \cdot n}$$

How can we analyze the distance?

$$g, g' \in \mathcal{C}_{AEL} \subseteq (\Sigma_{in}^d)^R$$

$$g \neq g'$$

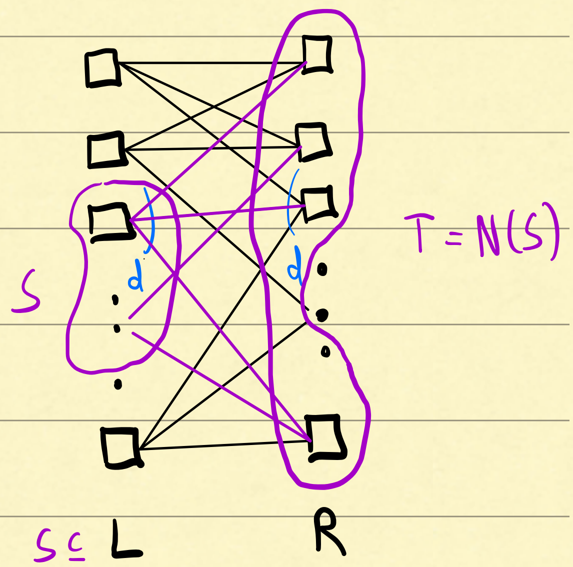
$$x, x' \in \mathcal{C}_{out}$$

$$\Delta(x,x') \geq \delta_{out}$$

$$\Delta(g,g') \geq |T|/n$$

$i \in S$ iff $x_i \neq x'_i$

$$|S| \geq \delta_{out} \cdot n$$



$$\boxed{\left|E(S,T) - \frac{d}{n}|S||T|\right| \leq \lambda d n}$$

$$\delta_{in} d |S| \leq E(S,T) \leq \frac{d}{n}|S||T| + \lambda d n$$

$$\Downarrow$$

$$\delta_{in} - \frac{\lambda n}{|S|} \leq \frac{|T|}{n} \qquad \boxed{|S| \geq \delta_{out} \cdot n}$$

$$\Downarrow$$

$$\delta_{in} - \lambda/\delta_{out} \leq \frac{|T|}{n} \leq \Delta(g,g')$$

How can we analyze the distance?

$g, g' \in \mathcal{C}_{AEL} \subseteq (\Sigma_{in}^{d})^{R}$

$g \neq g'$

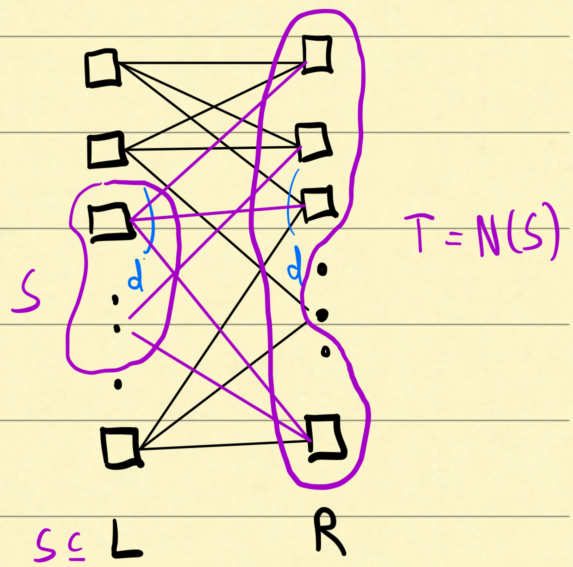
$x, x' \in \mathcal{C}_{out}$

$\Delta(x, x') \geq \delta_{out}$

$\Delta(g, g') \geq |T|/n$

$i \in S$ iff $x_i \neq x_i'$

$|S| \geq \delta_{out} \cdot n$



$$\boxed{\left|E(S,T) - \frac{d}{n}|S||T|\right| \leq \lambda d n}$$

$$\delta_{in} d |S| \leq E(S,T) \leq \frac{d}{n}|S||T| + \lambda d n$$

$$\Downarrow$$

$$\delta_{in} - \lambda \frac{n}{|S|} \leq \frac{|T|}{n} \qquad \boxed{|S| \geq \delta_{out} \cdot n}$$

$$\Downarrow$$

$$\delta_{in} - \underbrace{\lambda / \delta_{out}}_{\approx 0} \leq \frac{|T|}{n} \leq \Delta(g, g')$$

AEL Parameters

$$z(e_{AEL}) \geq z_{in} - \overbrace{\lambda / z_{out}}^{\approx 0}$$

$$r(e_{AEL}) = r(e_{out})\, r(e_{in})$$

AEL Parameters

$$\delta(\mathcal{C}_{AEL}) \geq \delta_{in} - \left(\cancel{\frac{\lambda}{\delta_{out}}}\right)^{\approx 0}$$

$$r(\mathcal{C}_{AEL}) = r(\mathcal{C}_{out})\, r(\mathcal{C}_{in})$$

We can take $\mathcal{C}_{in}$ with $\delta_{in} \geq 1 - R$
$r(\mathcal{C}_{in}) \geq R$
and $\mathcal{C}_{out}$ with $r(\mathcal{C}_{out}) \geq 1 - \varepsilon$

AEL Parameters

$\lambda \to 0$

$$\delta(\mathcal{C}_{AEL}) \geq \delta_{in} - \cancel{\frac{\lambda}{\delta_{out}}}^{\approx 0}$$

$$r(\mathcal{C}_{AEL}) = r(\mathcal{C}_{out})\, r(\mathcal{C}_{in})$$

We can take $\mathcal{C}_{in}$ with $\delta_{in} \geq 1-R$
$r(\mathcal{C}_{in}) \geq R$
and $\mathcal{C}_{out}$ with $r(\mathcal{C}_{out}) \geq 1-\varepsilon$

$$\boxed{\begin{aligned} \delta(\mathcal{C}_{AEL}) &\geq 1-R-\varepsilon \\ r(\mathcal{C}_{AEL}) &\geq R-\varepsilon \end{aligned}}$$

near the Singleton bound

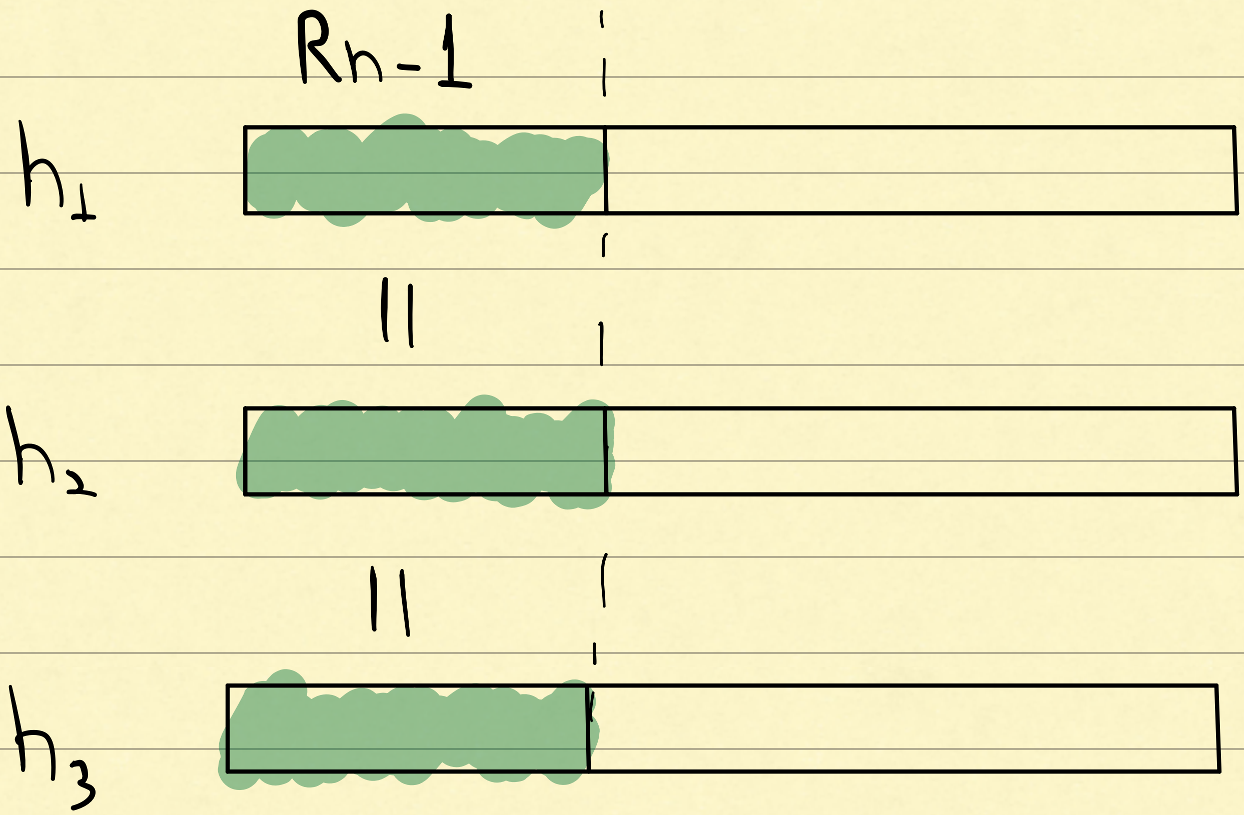
Is there a generalization of the Singleton bound for list decoding?

Generalized Singleton Bound

Let $\mathcal{C} \subseteq \Sigma^n$ of rate $R \in [0,1]$

$(|\mathcal{C}| = |\Sigma|^{Rn})$

$h_1, h_2, h_3 \in \mathcal{C}$ (distinct codewords)

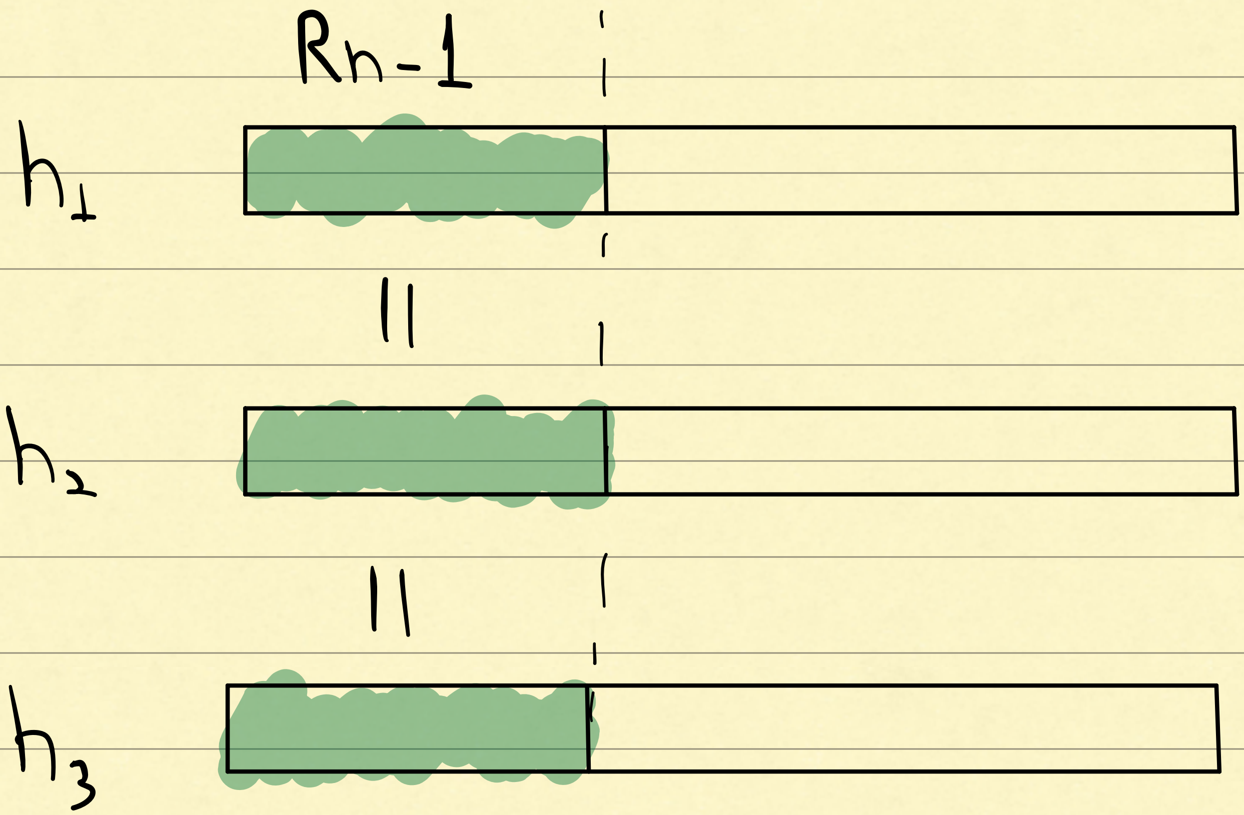


Generalized Singleton Bound

Let $\mathcal{C} \subseteq \Sigma^n$ of rate $R \in [0,1]$

$(|\mathcal{C}| = |\Sigma|^{Rn})$

$h_1, h_2, h_3 \in \mathcal{C}$ (distinct codewords)



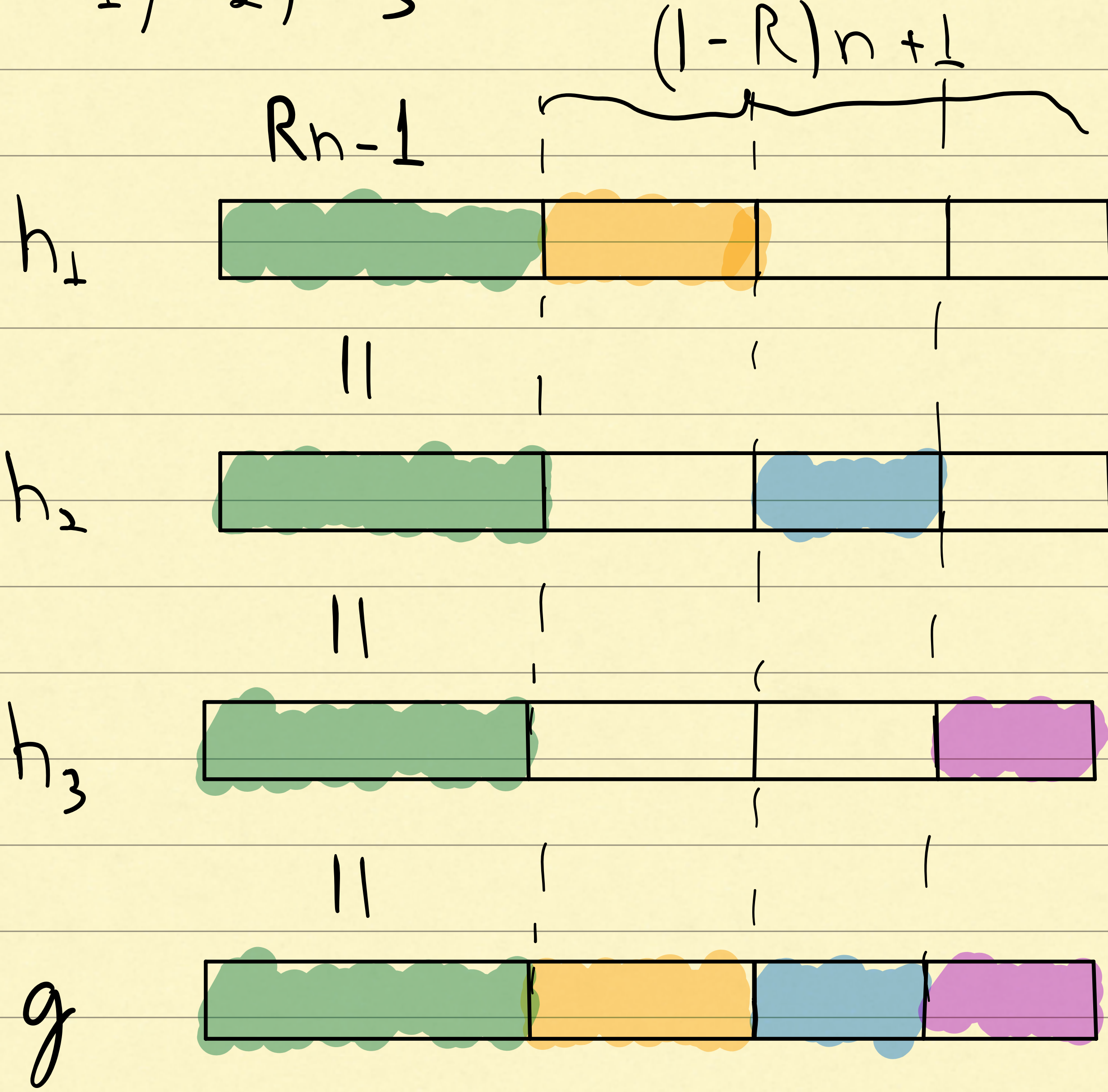
Can we construct a center $g \in \Sigma^n$ close to h_1, h_2, h_3?

Generalized Singleton Bound

Let $\mathcal{C} \subseteq \Sigma^n$ of rate $R \in [0,1]$

$(|\mathcal{C}| = |\Sigma|^{Rn})$

$h_1, h_2, h_3 \in \mathcal{C}$ (distinct codewords)



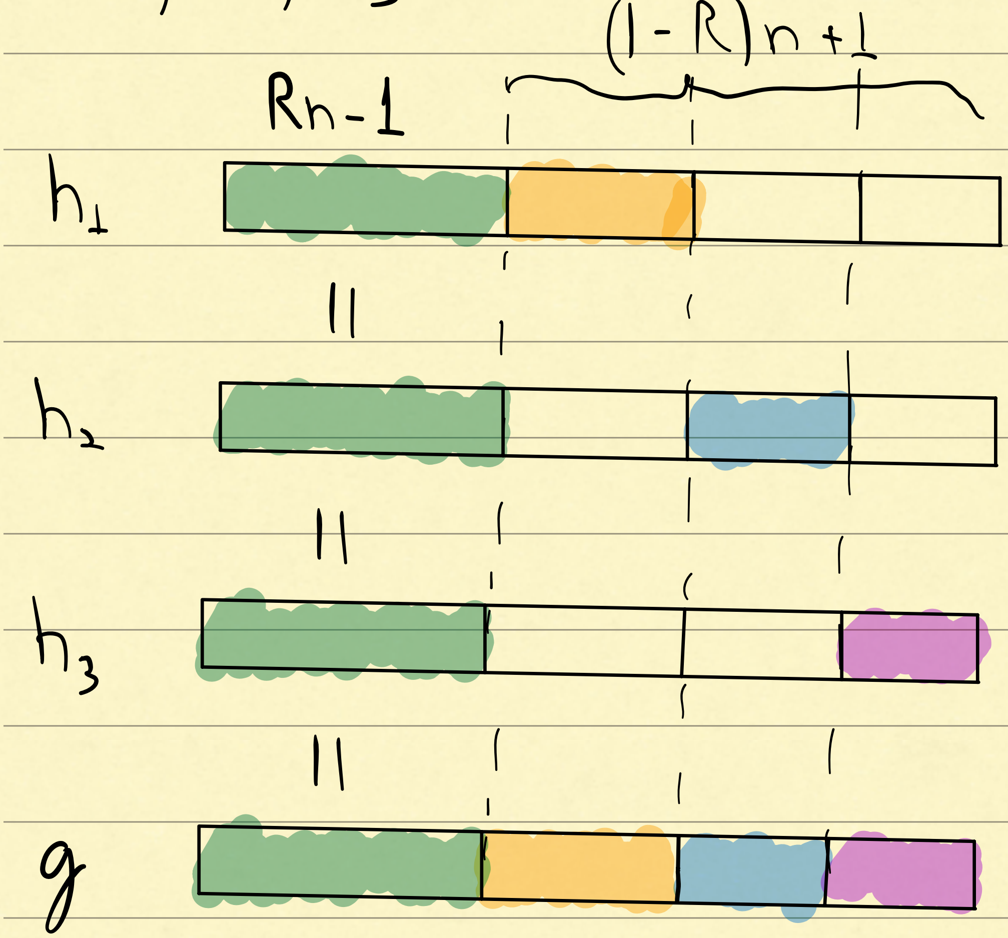
$h_1, h_2, h_3 \in \mathcal{B}\left(g, \frac{2}{3}\left(1-R+\frac{1}{n}\right)\right)$

Generalized Singleton Bound

Let $\mathcal{C} \subseteq \Sigma^n$ of rate $R \in [0,1]$

$$\left(|\mathcal{C}| = |\Sigma|^{Rn}\right)$$

$h_1, h_2, h_3 \in \mathcal{C}$ (distinct codewords)



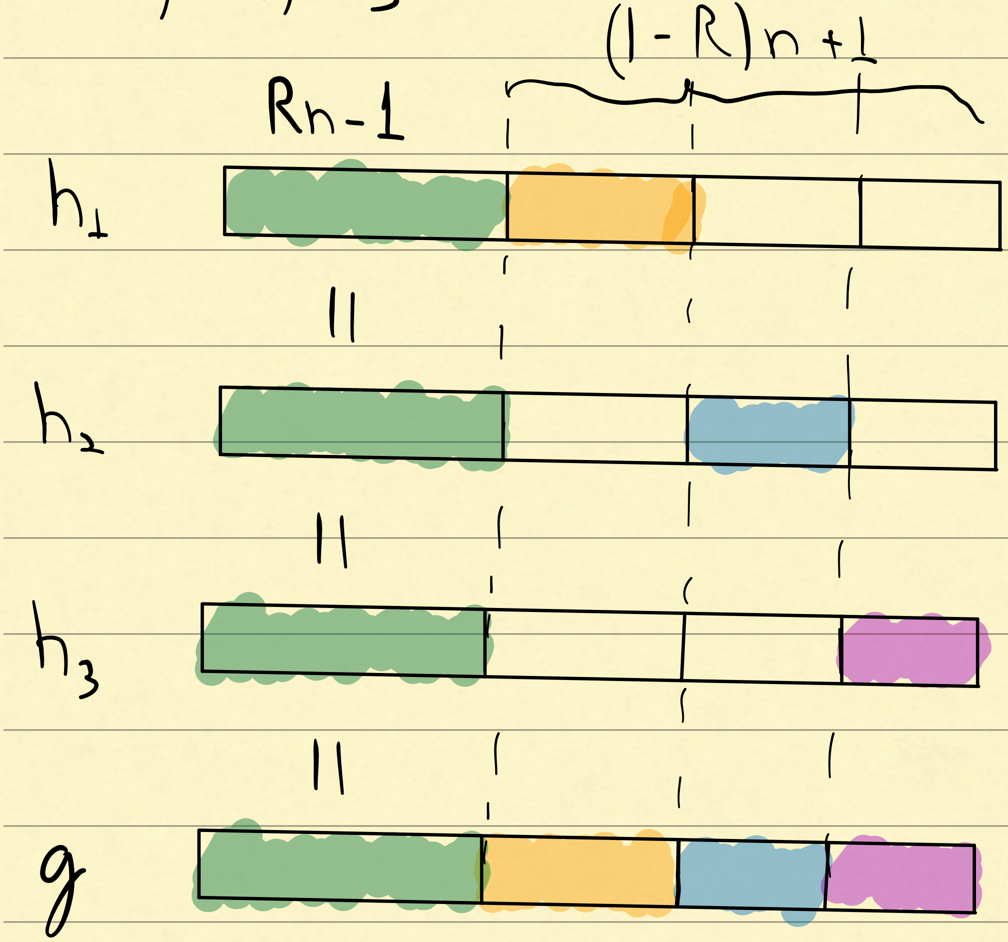
$$\left|\mathcal{L}\left(g, \tfrac{2}{3}(1-R)\right)\right| \geq 3$$

Generalized Singleton Bound

Let $\mathcal{C} \subseteq \Sigma^n$ of rate $R \in [0,1]$

$(|\mathcal{C}| = |\Sigma|^{Rn})$

$h_1, h_2, h_3 \in \mathcal{C}$ (distinct codewords)



Similar argument holds for K codewords

$$\left|\mathcal{L}\left(g, \frac{(K-1)}{K}(1-R)\right)\right| \geq K$$

Generalized Singleton Bound (GSB)

Thm [Shangguan-Tamo'20, Roth'22]

$$\left| \mathcal{L}\left(g, \frac{(K-1)}{K}(1-R)\right) \right| \geq K$$

$K=2$ in the original Singleton Bound

Generalized Singleton Bound (GSB)

Thm [Shangguan-Tamo'20, Roth'22]

$$\left|\mathcal{L}\left(\mathcal{C}, \frac{(K-1)}{K}(1-R)\right)\right| \geq K$$

$K=2$ is the original Singleton Bound

Cor: $\left|\mathcal{L}(\mathcal{C}, 1-R-\varepsilon)\right| \geq \Omega\left(\frac{1}{\varepsilon}\right)$

List sizes $\Omega\left(\frac{1}{\varepsilon}\right)$ are required for capacity

Moral:
Achieving list sizes of $O\left(\frac{1}{\varepsilon}\right)$ would be optimal

Folded Reed-Solomon (FRS)

List size bounds

- $n^{\frac{1}{\varepsilon}}$ [Guruswami-Rudra'06] [Guruswami'11]
- $\left(\frac{1}{\varepsilon}\right)^{\frac{1}{\varepsilon}}$ [Dvir-Lovett'12]
- $\frac{1}{\varepsilon^2}$ [Srivastava'25]
- $\frac{1}{\varepsilon}$ [Chen-Zhang'25] (approach GSB)

Folded Reed-Solomon (FRS)

Wonderful Properties

- List Decoding Capacity ✓
  [Guruswami-Rudra'06]
- Explicit ✓
- Efficient (List) Decoding ✓
- Approach GSB => optimal list sizes ✓
  [Chen-Zhang'25]

Major Drawbacks

- Large non-constant alphabet size
  ($\mathcal{C} \subseteq \mathbb{F}_q^n$ with $q = n^{1/\varepsilon^2}$) ↑
- No good "local" properties (LDPC) ↑

Dozens of Papers Studying List Decoding (Capacity) for Random Codes

- Powerful yardstick to analyze constructions ✓
- Beautiful Combinatorial/Probabilistic Ideas ✓

⋮

Dozens of Papers Studying List Decoding (Capacity) for Random Codes

- Powerful yardstick to analyze constructions ✓
- Beautiful Combinatorial/Probabilistic Ideas ✓

⋮

- Cannot efficiently certify $\Delta(C)$
- Cannot efficiently decode
- cannot be made explicit
- Cannot use in real applications

Dream Wish List of Coding Theory

We want a code with:

- Optimal Parameters
  (rate-vs-decoding radius)
- Explicit
- Alphabet as small as possible
- Efficient Encoding
- Efficient (list) Decoding
- Local Properties
  (LDPC, LTC, LDC, LCC, etc)

Our Results

Thm [J-Mittal-Srivastava-Tulsiani'25]

New family of codes

- Explicit
- Achieve List Decoding Capacity
- Over constant size alphabet!
- Optimal List Sizes (approach GSB)!
- "Efficient" List Decoding up to Capacity (Sum-of-Squares based)
- Local Properties: LDPC!

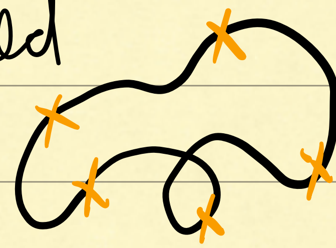
Known codes satisfying

[Guo-Ron-Zewi '20]

- Explicit ✓
- Achieve List Decoding Capacity ✓
- Over <u>constant</u> size alphabet ✓

are based on sophisticated

<u>Algebraic Geometry</u>



and still have the <u>drawbacks</u>:

- List sizes are far from optimal ∧
- Do not approach GSB ∧

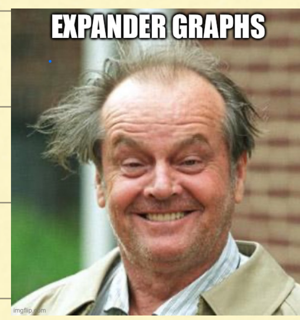
Our Results

Thm [J-Mittal-Srivastava-Tulsiani'25]

New family of codes

- Explicit
- Achieve List Decoding Capacity
- Over constant size alphabet!
- Optimal List Sizes (approach GSB)!
- "Efficient" List Decoding up to Capacity
  (Sum-of-Squares based)
- Local Properties: LDPC!

Based on Expander Graphs
(AEL for list decoding)



- Simple Construction
- Only uses Combinatorics
  (No need of sophisticated algebra)

Our Codes

Local-to-Global (using AEL)

Start with a constant size C_{in} approaching the generalized Singleton Bound* (GRS)

$$\Big\Downarrow \text{ obtain via AEL}$$

Global Code approaching the GRS

Our Codes

Local-to-Global (using AEL)

Start with a constant size $\mathcal{C}_{in}$ approaching the generalized Singleton Bound* (GRS)

$$\Big\Downarrow \text{ obtain via AEL}$$

Global Code approaching the GRS

(1) $\mathcal{C}_{in}$ can be a random linear code of $O(1)$ size

or

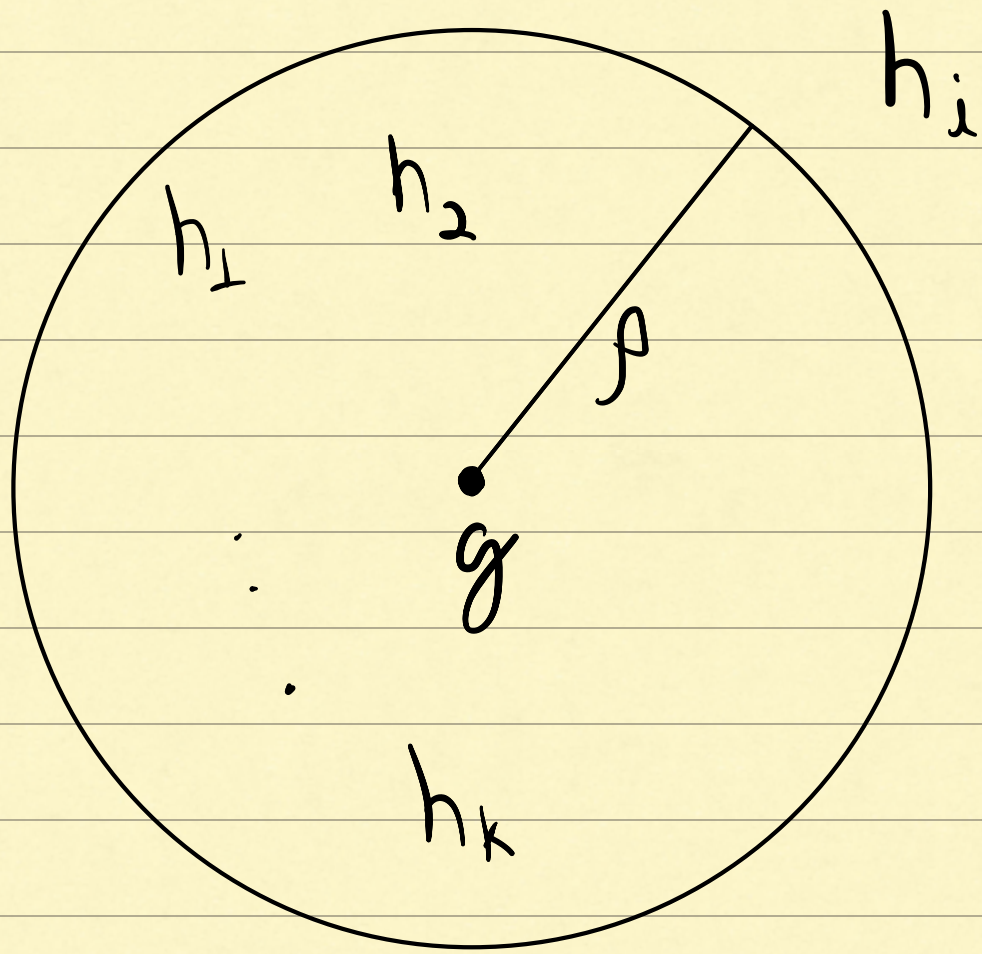
(2) $\mathcal{C}_{in}$ can be an explicit FRS of $O(1)$ size

List Decoding

Def: $\mathcal{C} \subseteq \Sigma^n$ is $(\rho, K-1)$-list decodable if

$\forall\, g \in \Sigma^n,\ h_1, \dots, h_K \in \mathcal{C}$

$$\max_{i \in [K]} \Delta(g, h_i) > \rho$$



"Not all $h_1, \dots, h_K$ can be in $B(g, \rho)$"

Average Radius List Decoding

(stronger than list decoding)

Def: $\mathcal{C} \subseteq \Sigma^n$ is $(\rho, K-1)$-list decodable if

$$\forall\, g \in \Sigma^n,\ h_1, \ldots, h_K \in \mathcal{C} \text{ (distinct)}$$

$$\mathop{\mathbb{E}}_{i \in [K]} \Delta(g, h_i) > \rho$$

Average Radius List Decoding

(stronger than list decoding)

Def: $\mathcal{C} \subseteq \Sigma^n$ is $(\rho, K-1)$-list decodable if

$\forall\, g \in \Sigma^n,\ h_1, \dots, h_K \in \mathcal{C}$ (distinct)

$$\mathbb{E}_{i \in [K]} \Delta(g, h_i) > \rho$$

ε-relaxed average GSB

$$\mathbb{E}_{i \in [K]} \Delta(g, h_i) \geq \frac{(K-1)}{K}(1-R-\varepsilon)$$

$$\Downarrow$$

$$\sum_{i \in [K]} \Delta(g, h_i) \geq (K-1)(1-R-\varepsilon)$$

Glimpse of The Proof

$$\mathrm{AEL}(G, \mathcal{C}_{out}, \mathcal{C}_{in})$$

$\llcorner$ ε-relaxed GSB

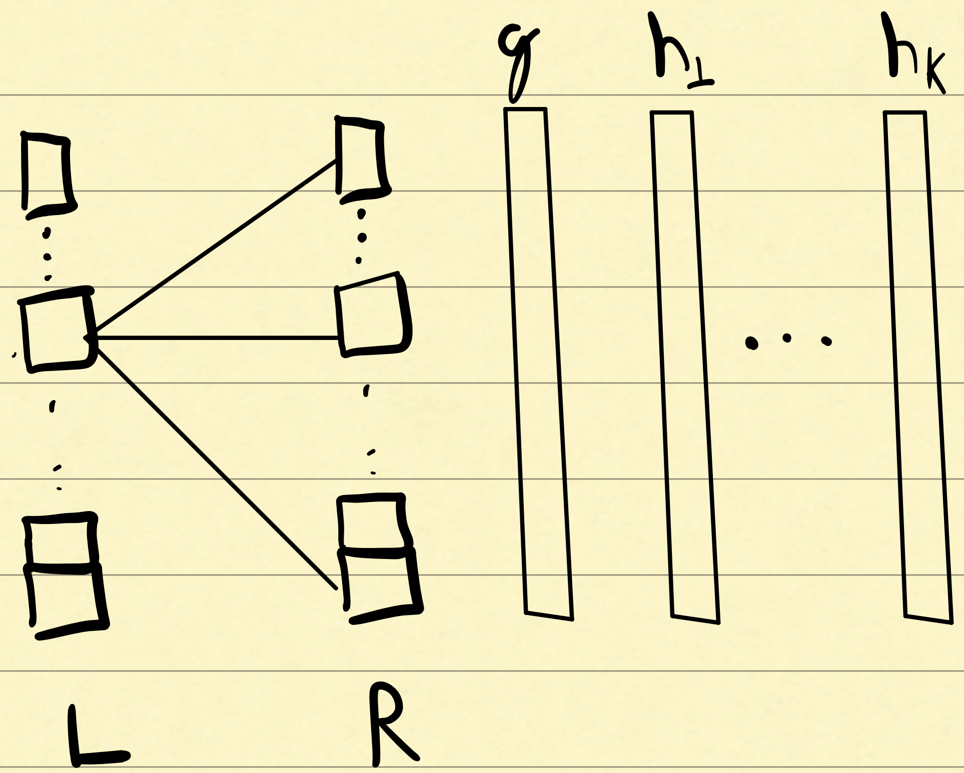
$\forall f \in \Sigma_{in}^{d}$, $\underbrace{f_1, \dots, f_k}_{\text{distinct}} \in \mathcal{C}_{in}$, $\sum_{i \in [k]} \Delta(f, f_i) \geq (k-1)(1-R-\varepsilon)$

$g \in (\Sigma_{in}^{d})^{n}$, $\underbrace{h_1, \dots, h_k}_{\text{distinct}} \in \mathcal{C}_{AEL}$

Local view from ℓth vertex

$g_\ell, \underbrace{f_{\ell,1}, \dots, f_{\ell,k}}$

$g_\ell \in \Sigma_{in}^{d}$, $\quad f_{\ell,1}, \dots, f_{\ell,k} \in \mathcal{C}_{in}$



Glimpse of The Proof

$\mathrm{AEL}(G, \mathcal{C}_{out}, \mathcal{C}_{in})$

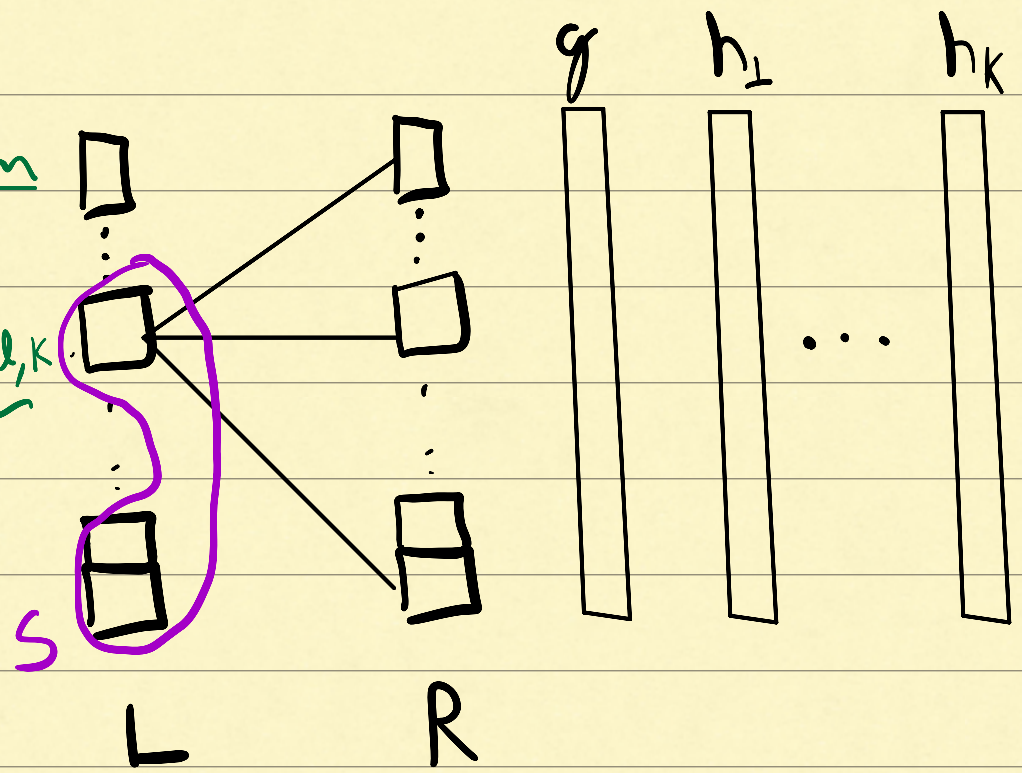
$\quad\llcorner$ ε-relaxed GSB

$\forall f \in \Sigma_{in}^d,\ \underbrace{f_1, \dots, f_K}_{\text{distinct}} \in \mathcal{C}_{in},\ \sum_{i \in [K]} \Delta(f, f_i) \geq (K-1)(1-R-\varepsilon)$

$g \in (\Sigma_{in}^d)^n,\ \underbrace{h_1, \dots, h_K}_{\text{distinct}} \in \mathcal{C}_{AEL}$

Local view from
lth vertex

$\underset{\Sigma_{in}^d}{\underset{\Uparrow}{g_l}},\ \underset{\mathcal{C}_{in}}{\underset{\Uparrow}{\underbrace{f_{l,1}, \dots, f_{l,K}}}}$



$S = \{\, l \in [n] \mid f_{l,1}, \dots, f_{l,K} \text{ are not all equal} \,\}$

Glimpse of The Proof

$\mathrm{AEL}(\mathcal{G}, \mathcal{C}_{out}, \mathcal{C}_{in})$

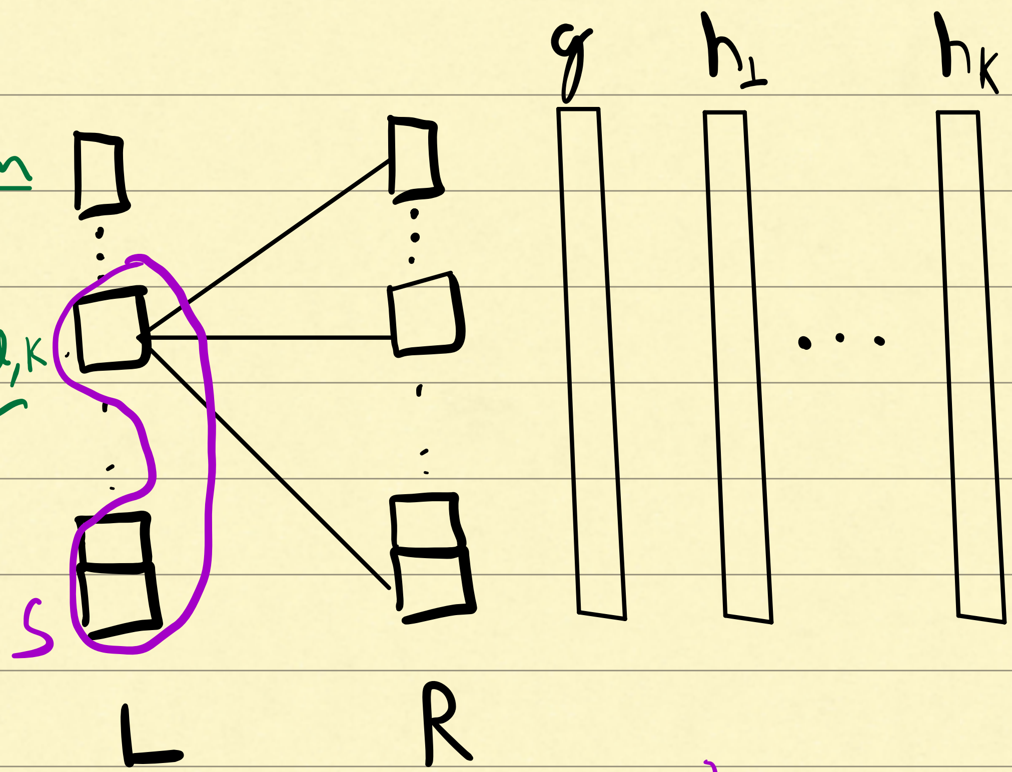
$\llcorner$ ε-relaxed GSB

$\forall f \in \Sigma_{in}^{d}, \underbrace{f_1, \ldots, f_k}_{\text{distinct}} \in \mathcal{C}_{in}, \sum_{i \in [k]} \Delta(f, f_i) \geq (k-1)(1-R-\varepsilon)$

$g \in (\Sigma_{in}^{d})^{n}, \underbrace{h_1, \ldots, h_k}_{\text{distinct}} \in \mathcal{C}_{AEL}$

Local view from lth vertex

$\underset{\Sigma_{in}^{d}}{\underset{\cap}{g_l}}, \underbrace{f_{l,1}, \ldots, f_{l,k}}_{\underset{\mathcal{C}_{in}}{\cap}}$



$S = \{ l \in [n] \mid f_{l,1}, \ldots, f_{l,k} \text{ are not all equal} \}$

$|S| \geq \delta_{out} \cdot n$

Glimpse of The Proof

$AEL(\mathcal{G}, \mathcal{C}_{out}, \mathcal{C}_{in})$

$\llcorner$ ε-relaxed GSB

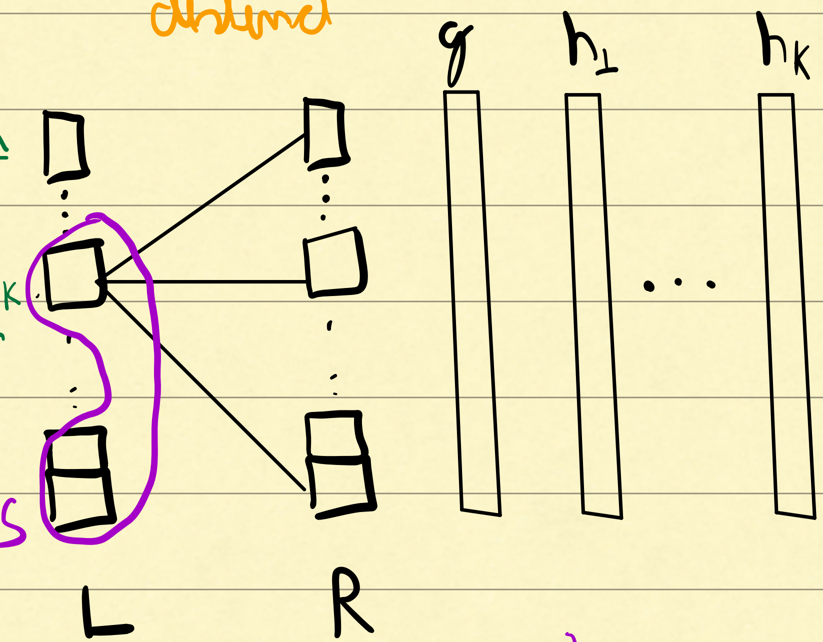
$\forall f \in \Sigma_{in}^d$, $\underbrace{f_1,\dots,f_K}_{\text{distinct}} \in \mathcal{C}_{in}$, $\sum_{i\in[K]} \Delta(f,f_i) \geq (K-1)(1-R-\varepsilon)$

$g \in (\Sigma_{in}^d)^n$, $\underbrace{h_1,\dots,h_K}_{\text{distinct}} \in \mathcal{C}_{AEL}$

Local view from ℓth vertex

$g_\ell, \underbrace{f_{\ell,1},\dots,f_{\ell,K}}$

$g_\ell \in \Sigma_{in}^d$, $f_{\ell,1},\dots,f_{\ell,K} \in \mathcal{C}_{in}$



$S = \{\ell \in [n] \mid f_{\ell,1},\dots,f_{\ell,K}$ are not all equal$\}$

$|S| \geq \delta_{out} \cdot n$

Easy Case: $\forall \ell \in S, |\{f_{\ell,1},\dots,f_{\ell,K}\}| = K$

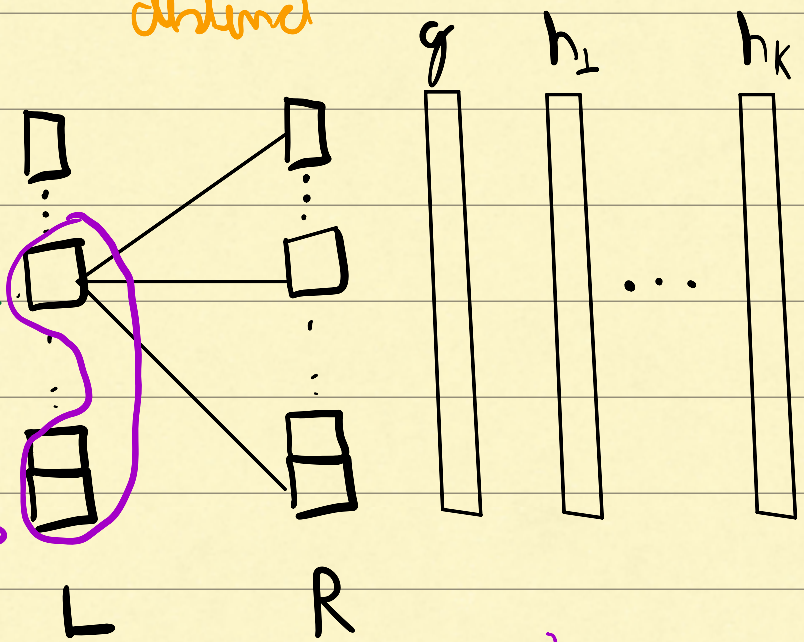
Glimpse of The Proof

$$\mathrm{AEL}(\mathcal{G}, \mathcal{C}_{out}, \mathcal{C}_{in})$$

$\mathcal{C}_{in}$: ε-relaxed GSB

$\forall f \in \Sigma_{in}^{d}$, $\underbrace{f_1, \dots, f_K}_{\text{distinct}} \in \mathcal{C}_{in}$, $\displaystyle\sum_{i \in [K]} \Delta(f, f_i) \geq (K-1)(1-R-\varepsilon)$

$g \in (\Sigma_{in}^{d})^{n}$, $\underbrace{h_1, \dots, h_K}_{\text{distinct}} \in \mathcal{C}_{AEL}$



Local view from ℓth vertex

$\underset{\Sigma_{in}^{d}}{\overset{}{g_\ell}}, \underbrace{f_{\ell,1}, \dots, f_{\ell,K}}_{\in\ \mathcal{C}_{in}}$

($g_\ell \in \Sigma_{in}^{d}$, $f_{\ell,1}, \dots, f_{\ell,K} \in \mathcal{C}_{in}$)

$S = \{\ell \in [n] \mid f_{\ell,1}, \dots, f_{\ell,K}$ are not all equal$\}$ $\quad |S| \geq \delta_{out} \cdot n$

Easy Case: $\forall \ell \in S, |\{f_{\ell,1}, \dots, f_{\ell,K}\}| = K$

$$\Delta(g, h_j) \gtrsim \mathop{\mathbb{E}}_{\ell \in S} \Delta(g_\ell, f_{\ell,j}) \quad \forall j \in [K]$$

Glimpse of The Proof

$$AEL(G, \mathcal{C}_{out}, \mathcal{C}_{in})$$

$\mathcal{C}_{in}$: ε-relaxed GSB

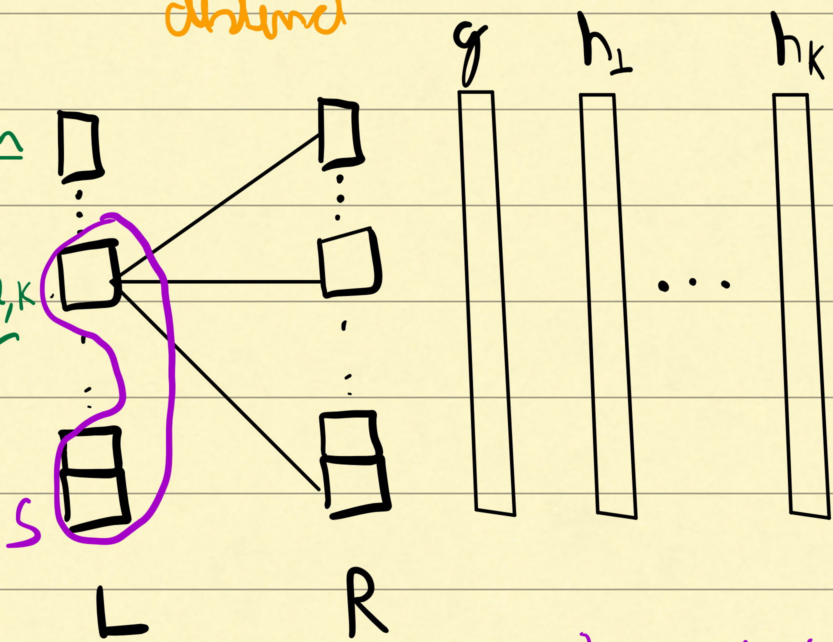
$$\forall f \in \Sigma_{in}^d, \underbrace{f_1, \dots, f_K}_{\text{distinct}} \in \mathcal{C}_{in}, \sum_{i \in [K]} \Delta(f, f_i) \geq (K-1)(1-R-\varepsilon)$$

$$g \in (\Sigma_{in}^d)^n, \underbrace{h_1, \dots, h_K}_{\text{distinct}} \in \mathcal{C}_{AEL}$$

Local view from ℓth vertex

$g_\ell, \underbrace{f_{\ell,1}, \dots, f_{\ell,K}}$

$g_\ell \in \Sigma_{in}^d$, $f_{\ell,1}, \dots, f_{\ell,K} \in \mathcal{C}_{in}$



$S = \{\ell \in [n] \mid f_{\ell,1}, \dots, f_{\ell,K} \text{ are not all equal}\}$ $\quad |S| \geq \delta_{out} \cdot n$

Easy Case: $\forall \ell \in S, |\{f_{\ell,1}, \dots, f_{\ell,K}\}| = K$

$$\sum_{i \in [K]} \Delta(g, h_i) \gtrsim \mathbb{E}_{\ell \in S} \sum_{i \in [K]} \Delta(g_\ell, f_{\ell,i})$$

Glimpse of The Proof

$\mathrm{AEL}(G, \mathcal{C}_{out}, \mathcal{C}_{in})$

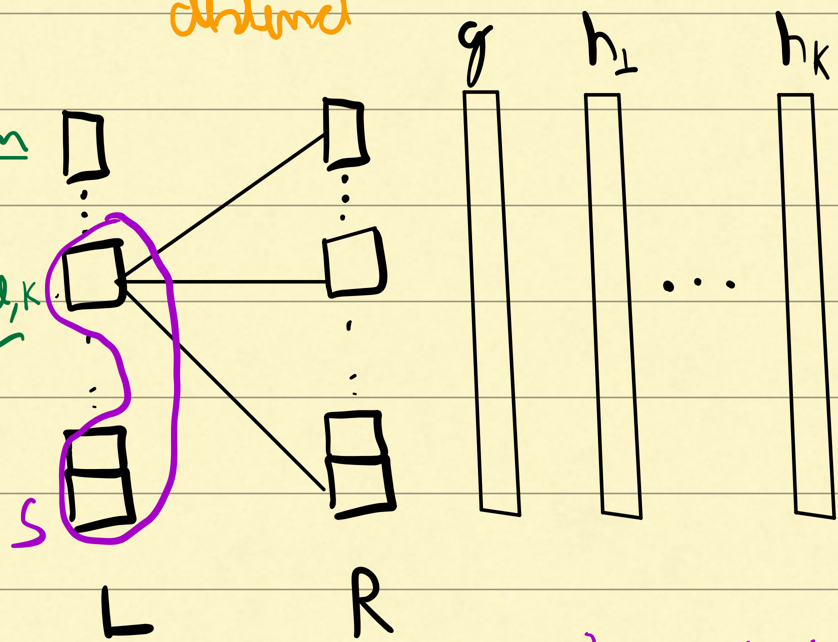
$\llcorner$ ε-relaxed GSB

$\forall f \in \Sigma_{in}^{d},\ \underbrace{f_1, \ldots, f_k}_{\text{distinct}} \in \mathcal{C}_{in},\ \sum_{i \in [k]} \Delta(f, f_i) \geq (k-1)(1-R-\varepsilon)$

$g \in (\Sigma_{in}^{d})^{n},\ \underbrace{h_1, \ldots, h_k}_{\text{distinct}} \in \mathcal{C}_{AEL}$

Local view from lth vertex

$\underset{\Sigma_{in}^{d}}{\overset{\cap}{g_l}},\ \underbrace{f_{l,1}, \ldots, f_{l,k}}_{\overset{\cap}{\mathcal{C}_{in}}}$



$S = \{ l \in [n] \mid f_{l,1}, \ldots, f_{l,k} \text{ are not all equal} \}$ $\quad |S| \geq \delta_{out} \cdot n$

Easy Case: $\forall l \in S,\ |\{f_{l,1}, \ldots, f_{l,k}\}| = k$

$$\sum_{i \in [k]} \Delta(g, h_i) \gtrsim (k-1)(1-R-\varepsilon)$$

$\blacksquare$

Open Question Galore

- More efficient Decoding
- Implications to other pseudorandom constructions (e.g., extractors, lossless expanders)
- Tight(er) alphabet size
- Quantum Capacity
- List Recovery / Other models

Open Problems

Thank you!