

LOWER BOUNDS FOR 2- QUERY LOCALLY CORRECTABLE CODES OVER FINITE FIELDS

Arnab Bhattacharyya

Zeev Dvir

Subhangi Saraf

Amir Shpilka

Overview of talk

- Definitions
 - Locally correctable codes (LCCs)
- Connection to other problems
 - Locally Decodable Codes
 - Extensions of Sylvester-Gallai theorem
 - Extensions of Beck's theorem
 - Rank of design matrices
 - Additive Combinatorics
- Some proofs

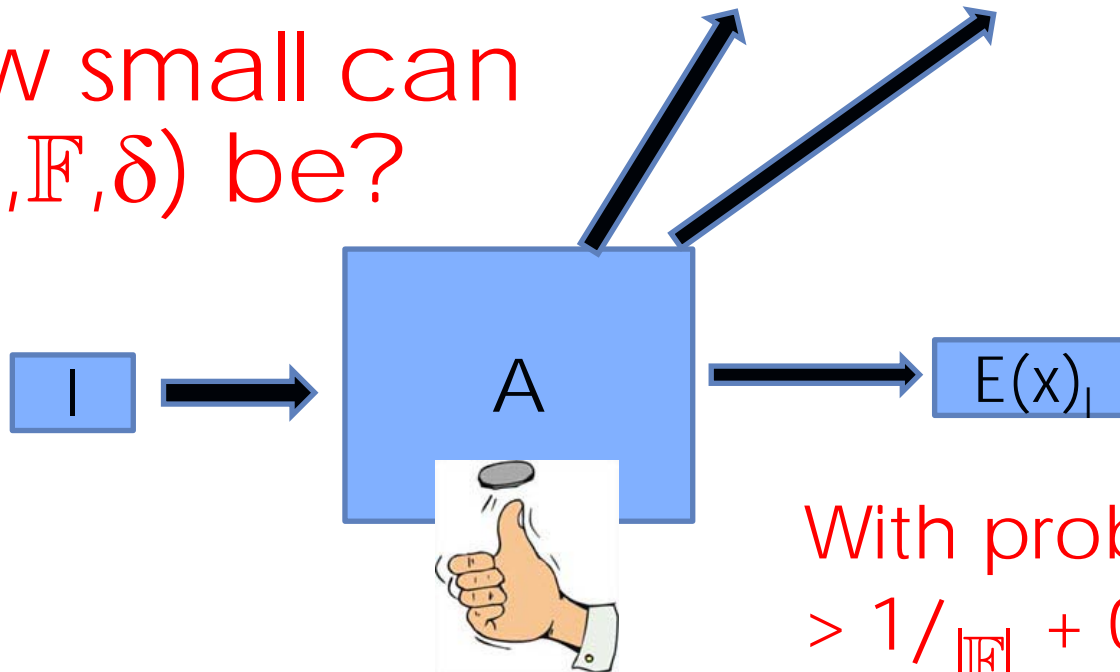
Conventions

- All codes are linear
- All codes over a finite field \mathbb{F}
(think $\mathbb{F} = \mathbb{F}_p$)
- k = message length = dimension of code
- n = block length

$(2, \delta)$ -Locally correctable codes I



Q: how small can $n = n(k, \mathbb{F}, \delta)$ be?



With probability $> 1/|\mathbb{F}| + 0.1$

$(2, \delta)$ -Locally correctable codes I



- Q:
n=
- Noise is adversarial
 - Algorithm should work for every message x
 - The error probability is on the randomness of the decoder and not because of input/noise distributions!

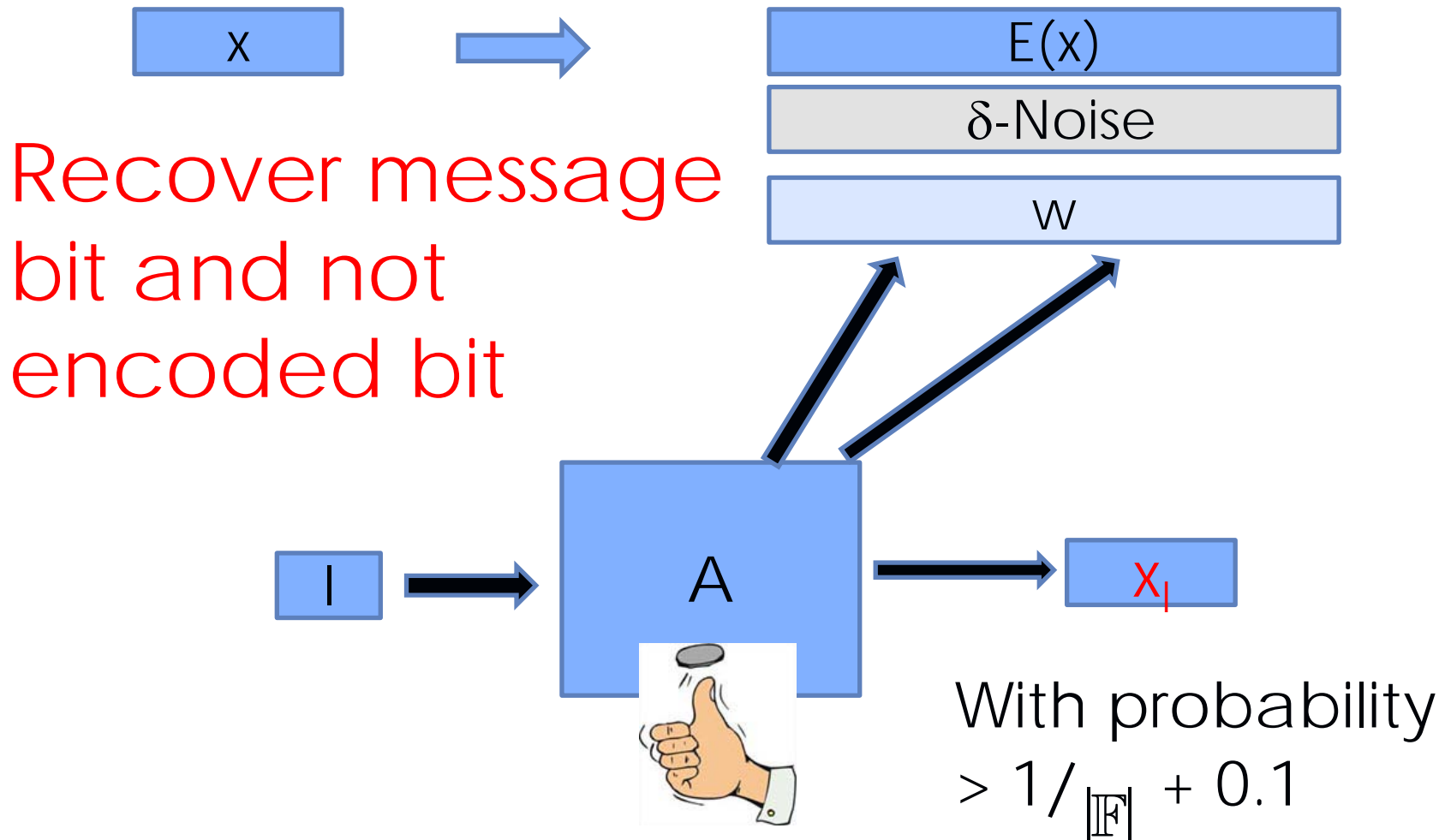


With probability
 $> 1/|\mathbb{F}| + 0.1$

Related to...

- Locally Decodable codes
- Quantitative versions of
 - Sylvester-Gallai theorem
 - Beck theorem
- Rank of design matrices
- Additive combinatorics
 - If $|A-A| \sim |A|$, how close is A to a vector space?

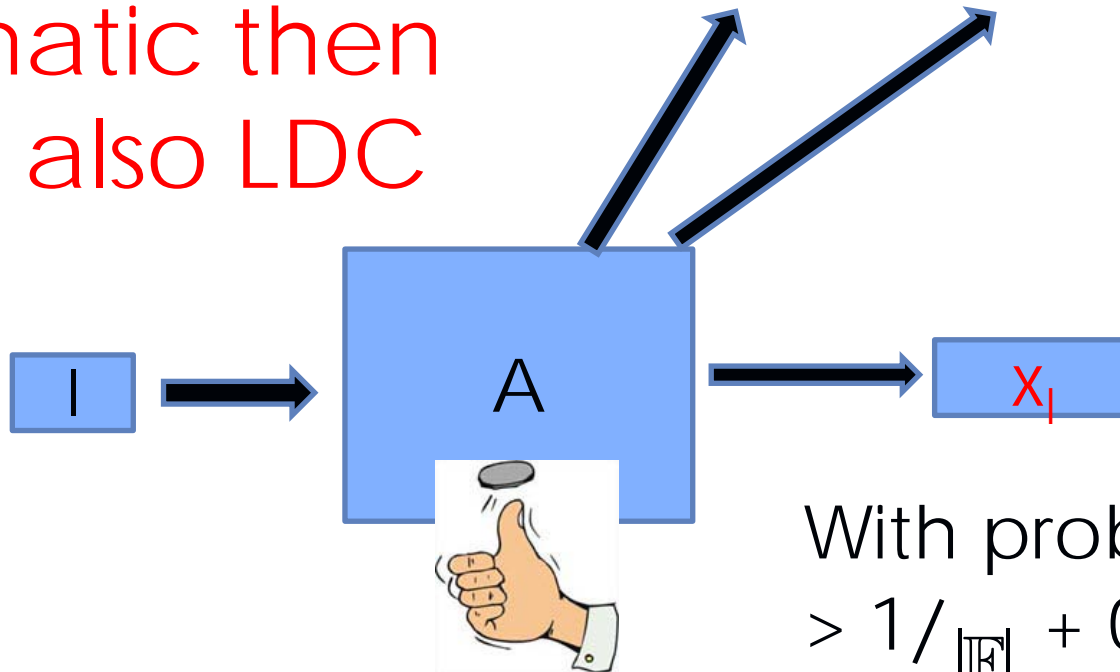
$(2, \delta)$ -Locally decodable codes



$(2, \delta)$ -Locally decodable codes



If the code is systematic then LCC is also LDC



With probability $> 1/|\mathbb{F}| + 0.1$

Example

- “Hadamard” over \mathbb{F} :
 $x \rightarrow (\langle x, v \rangle) : v \in \{0, 1\}^k$
- **LDC for any \mathbb{F}** : for $l \in [k]$ pick random v s.t. $v_l = 1$. Return $\langle x, v \rangle - \langle x, v - e_l \rangle$
- **LCC over \mathbb{F}_2** : For $u \in \{0, 1\}^k$ pick random v . Return $\langle x, v \rangle \oplus \langle x, v \oplus u \rangle$
- What happens over \mathbb{F}_p ?
Can we get same parameters as \mathbb{F}_2 ?

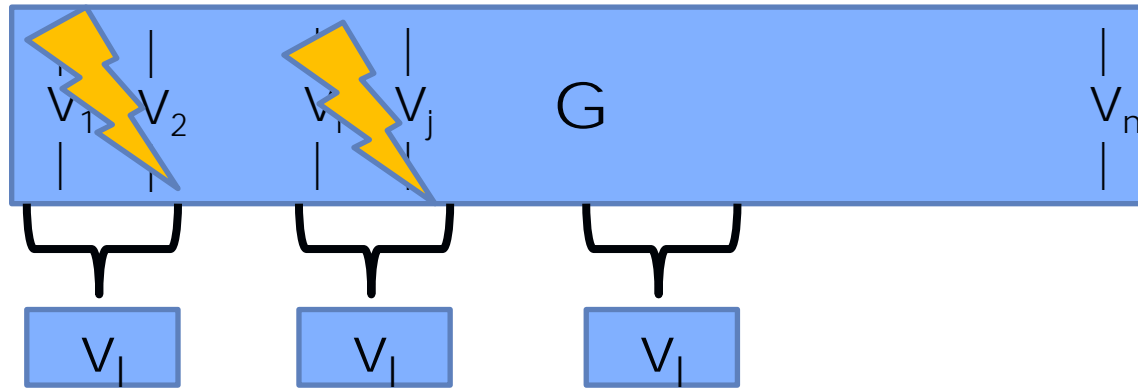
Recovering via Matching I

$$E(x) = x \begin{array}{c|c|c|c} & v_1 & v_2 & \dots & v_n \\ \hline & | & | & & | \\ & | & | & & | \\ \hline & & & G & \\ \hline & | & | & & | \\ & | & | & & | \end{array}$$

$$E(x) = (\langle x, v_1 \rangle, \langle x, v_2 \rangle, \dots, \langle x, v_n \rangle)$$

If $\langle x, v_i \rangle$ can be recovered from $\langle x, v_i \rangle, \langle x, v_j \rangle$
then $v_i \in \text{span}\{v_i, v_j\}$

Recovering via Matching II



Each v_1 is spanned by δn disjoint pairs
Each v_1 has a matching of size δn

$(2, \delta)$ -Locally correctable codes II

- List of n vectors (v_1, \dots, v_n) in \mathbb{F}^k
- Each v_i has a matching M_i on $[n]$ s.t.
 - $|M_i| \geq \delta n$
 - if $(i, j) \in M_i$ then $v_i \in \text{span}\{v_i, v_j\}$

Main Questions:

- Construct such codes
- How large should $n = n(k, \mathbb{F}, \delta)$ be?

LDCs vs. LCCs

- Each LCC is LDC
- Over any \mathbb{F} LDCs have length $n=2^{\Omega(\delta k)}$
- Tight because of Hadamard
- For LCCs this is tight only for \mathbb{F}_2
- What happens over other fields?
- Best construction: generalized Hadamard
 $x \rightarrow (\langle x, v \rangle) : v \in (\mathbb{F}_p)^k$
 $n = p^k$
- **Main question:** what is the "truth", p^k or 2^k ?
(is LCC a stronger requirement than LDC?)

Main Result

□ Theorem: Let (v_1, \dots, v_n) be a $(2, \delta)$ -LCC in \mathbb{F}^k . Then

$$k \leq \text{poly}(p/\delta) + (2/\delta)\log_p(n)$$

□ In other words, $n = p^{\Omega(\delta k)}$

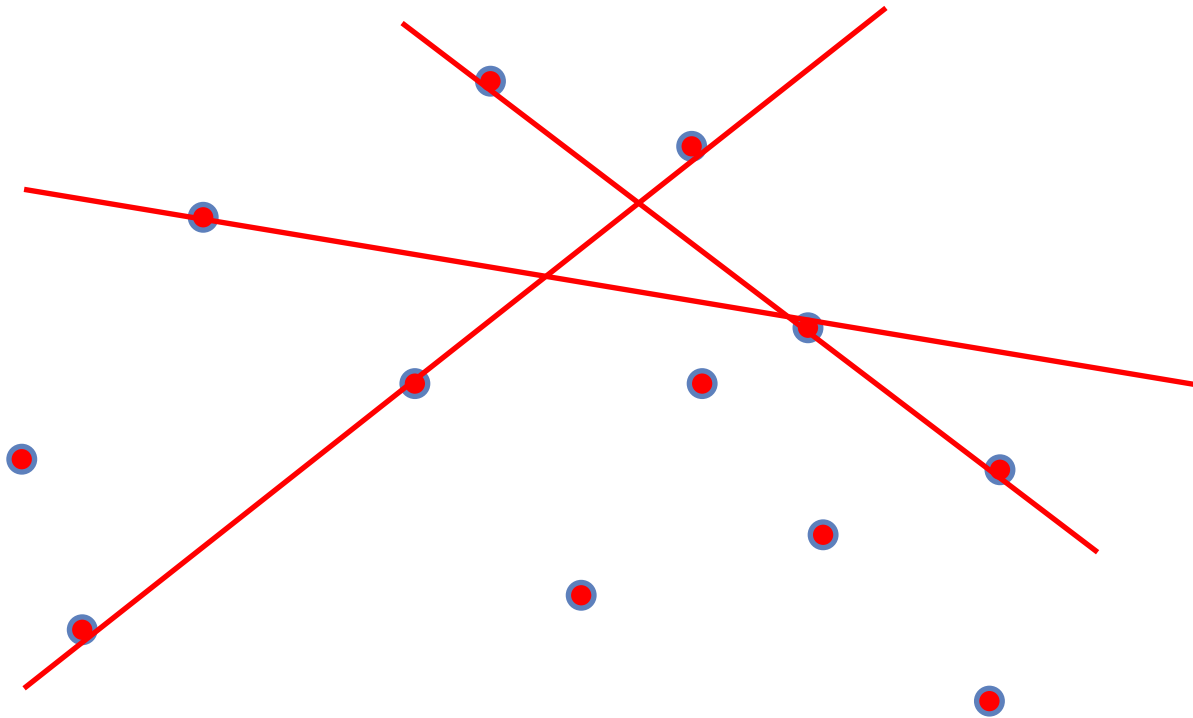
□ \Rightarrow LCCs need to be longer than LDCs

Related to...

- Locally Decodable codes
- Quantitative versions of
 - Sylvester-Gallai theorem
 - Beck theorem
- Rank of design matrices
- Additive combinatorics
 - If $|A-A| \sim |A|$, how close is A to a vector space?

Sylvester-Gallai theorem

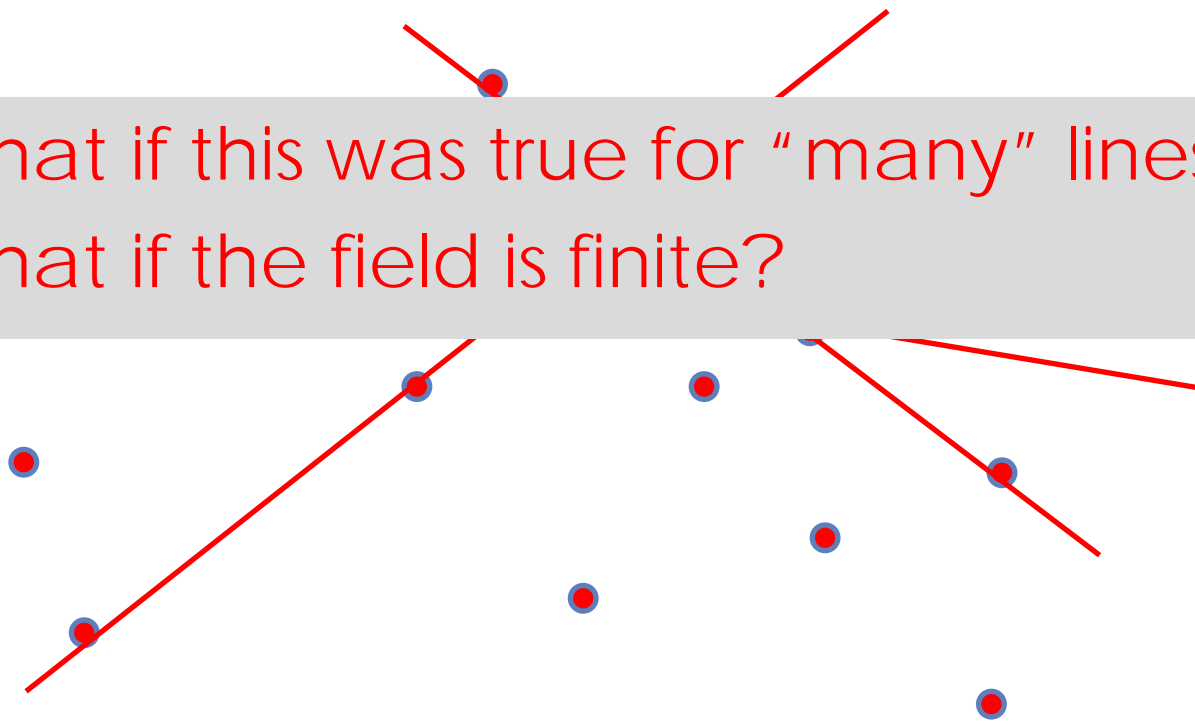
Theorem [S-G]: n points in \mathbb{R}^k such that any line through any 2 of them contains a 3rd point, must be co-linear



Sylvester-Gallai theorem

Theorem [S-G]: n points in \mathbb{R}^k such that any line through any 2 of them contains a 3rd point, must be co-linear

- ▣ What if this was true for “many” lines?
- ▣ What if the field is finite?



Sylvester-Gallai theorem

- Theorem [S-G]: n points in \mathbb{R}^k such that **any line** through any 2 of them contains a 3rd point, must be **co-linear**
- Theorem[B-D-W-Y]: n points in \mathbb{R}^k such that **δn lines** through any one of them contain ≥ 3 points, have dimension **$k \leq 1/\delta^2$**
- Theorem: n points in \mathbb{F}^k such that **δn lines** through any one of them contain ≥ 3 points, have dimension **$k \leq O(\log_p(n)/\delta)$**

Proof of extended S-G thm

- **Theorem:** n points in \mathbb{F}^k such that δn lines through any one of them contain ≥ 3 points, have dimension $k \leq O(\log_p(n)/\delta)$
- **Proof:** Let (v_1, \dots, v_n) in \mathbb{F}^k be those points
- $\forall v_i \exists$ "many" pairs co-linear with it
 \Rightarrow each v_i has a matching M_i on $[n]$ s.t.
 - $|M_i| = \Omega(\delta n)$
 - if $(i, j) \in M_i$ then $v_i \in \text{span}\{v_i, v_j\}$
- **Can use bounds on LCCs**

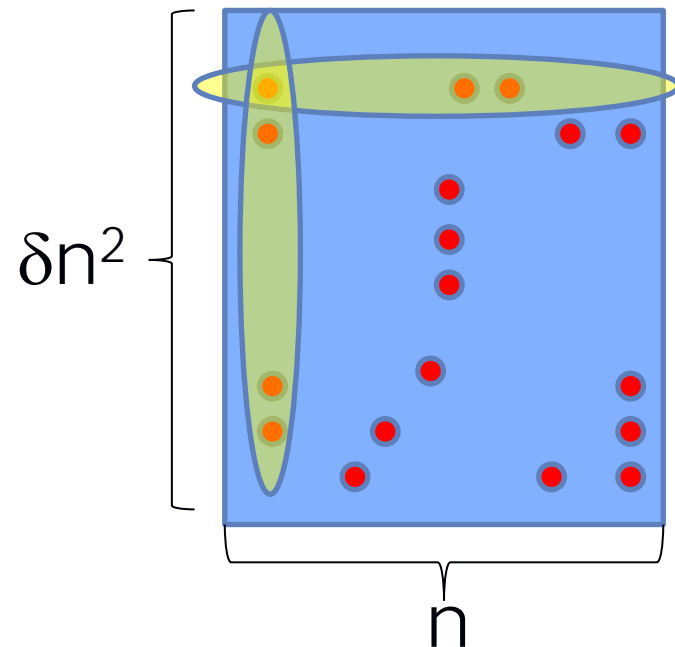
Beck's theorem

- **Theorem [Beck]:** $\exists \alpha, \beta > 0$ such that if n points in \mathbb{R}^2 determine only αn^2 lines then βn of them are colinear
- What if the field is finite?
- **Theorem:** If n points in \mathbb{F}^k determine only αn^2 lines then $\frac{1}{2}$ of them span a space of dimension $\leq O(\log_p(n)/\delta)$
($\alpha < 1/64, \delta = 1 - 8\sqrt{\alpha}$)

Rank of design matrices

- $(3, m, t)$ -matrix
- ≤ 3 1's in a row
- $\geq m$ 1's in a column
- Two columns share $\leq t$ 1's
- How small can the rank be?

Example: parity checks of 2-LCC



Theorem: rank of $(3, \delta n, t)$ -design is at least $n - O(t \log_p(n)/\delta)$

Related to...

- Locally Decodable codes
- Quantitative versions of
 - Sylvester-Galai theorem
 - Beck theorem
- Rank of design matrices
- Additive combinatorics
 - If $|A-A| \sim |A|$, how close is A to a vector space?

(n.w.l.o.g.) Simplifying assumptions on LCCs (t.c.b.r.)

Recall definition:

- List of n vectors (v_1, \dots, v_n) in \mathbb{F}^k
- Each v_i has a matching M_i on $[n]$ s.t.
 - $|M_i| \geq \delta n$
 - if $(i,j) \in M_i$ then $v_i \in \text{span}\{v_i, v_j\}$

Assume:

- no repetitions (all v_i distinct)
- $v_i \in \text{span}\{v_i, v_j\} \Rightarrow v_i = v_i - v_j$

Proof of lower bound on LCC

□ Theorem: Let (v_1, \dots, v_n) be a $(2, \delta)$ -LCC in \mathbb{F}^k . Then

$$k \leq \text{poly}(p/\delta) + (2/\delta)\log_p(n)$$

□ In other words, $n = p^{\Omega(\delta k)}$

Ruzsa's theorem

- ▣ Theorem [R]: $A \subseteq (\mathbb{F}_p)^k$ satisfy $|A-A| < c|A|$. Then,
 \exists subspace W s.t. $A \subseteq W$ and

$$|W| \leq \exp(c^4)|A|$$

$$\Rightarrow \dim(W) < O(c^4) + \log_p |A|$$

- ▣ I.e. A is "almost" a subspace

- ▣ Set $A = \{v_1, \dots, v_n\}$ (recall, no repetitions)

If we could apply Ruzsa's thm then

$$k = \dim(A) \leq \dim(W) < O(c^4) + \log_p |A| = O(c^4) + \log_p n$$

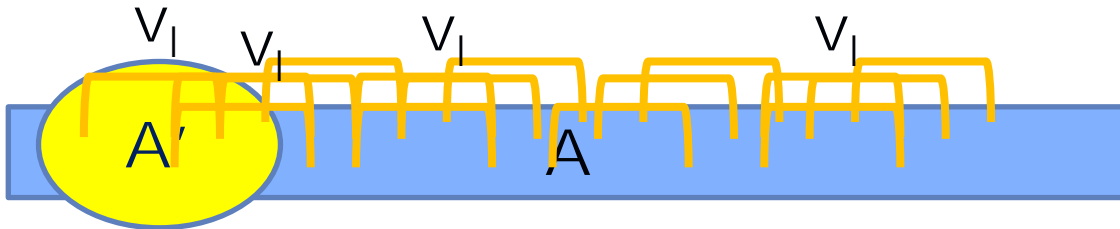
Ruzsa's theorem

- **Theorem [R]:** $A \subseteq (\mathbb{F}_p)^k$ satisfy $|A-A| < c|A|$. Then, \exists subspace W s.t. $A \subseteq W$ and
$$|W| \leq \exp(c^4)|A|$$
$$\Rightarrow \dim(W) < O(c^4) + \log_p |A|$$
- I.e. A is "almost" a subspace
- Set $A = \{v_1, \dots, v_n\}$ (recall, no repetitions)
- Each v_i is equal to $v_i - v_j$ for many pairs
- $\sim \delta n^2$ differences in $A-A$ belong to A
- However, Ruzsa demands that all differences belong to a small set

Proof of lower bound on LCC II

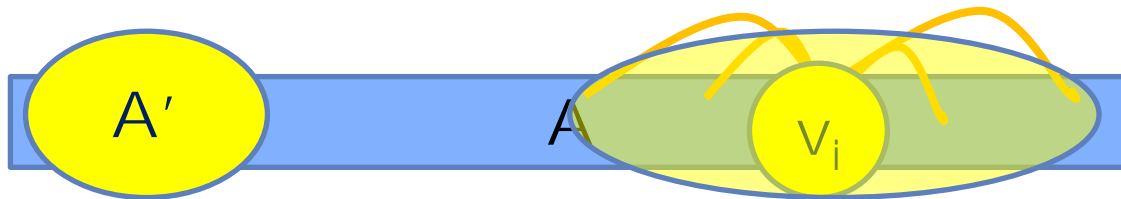
- **Step I:** use Balog-Gowers-Szemeredy lemma to find $A' \subseteq A$ such that
 - $|A'| > (\delta/p) \cdot |A|$ (A' is large)
 - $|A' - A'| < \text{poly}(p/\delta) \cdot |A'|$ (can apply Ruzsa)
- **Step II:** By Ruzsa, $\dim(A') < \log_p |A'|$
- **Step III:** Amplify A' to include $\geq \delta$ fraction of all points in A
- **Step IV:** Remove A' from A . Recurse.

Amplifying A'



- Each v_i in A' has M_i
- # of edges $> \delta n |A'| > |A'|^2$

Amplifying A'

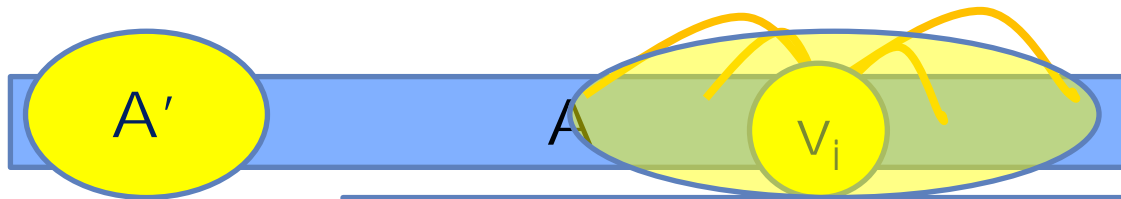


- Each v_i in A' has M_i
- # of edges $> \delta n |A'| > |A'|^2$
- $\exists v_i \in A \setminus A'$ of degree $\sim \text{poly}(p, \delta)n$
- $A' \leftarrow \text{span}(A' \cup \{v_i\})$.

Dim + 1. Size + $\text{poly}(p, \delta)n$

- Can repeat only $1/\text{poly}(p, \delta)$ times

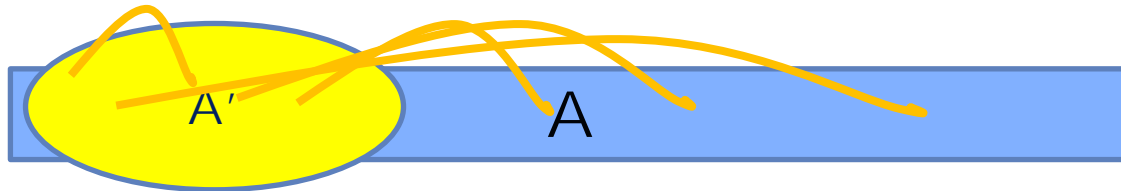
Amplifying A'



Edges labeled by $v_i \in A'$.
 With v_i we span its neighbors.

- Each
- # of e
- $\exists v_i \in$
- $A' \leftarrow \text{span}(A' \cup \{v_i\})$.
- Dim + 1. Size + $\text{poly}(p, \delta)n$
- Can repeat only $1/\text{poly}(p, \delta)$ times

Recursion



- Assume $A \setminus A'$ not $(2, \delta/2)$ -LCC
- $\exists v_i \in A \setminus A'$ with no good matching
- $\Rightarrow M_i$ has $\delta n/2$ edges touching A'
- $A' \leftarrow \text{span}(A' \cup \{v_i\})$. Dim + 1. Size + $\delta n/2$.
- Can repeat only $2/\delta$ times

Where is the cheat?

- Recall “(n.w.l.o.g.) Simplifying assumptions on LCCs (t.c.b.r.)”

Assume:

- no repetitions
- $v_l \in \text{span}\{v_i, v_j\} \Rightarrow v_l = v_i - v_j$
- Main (technical) lemma: can pass to a subcode with all multiplicities equal

Open problems

- LDC constructions with $O(1)$ queries of size $\exp(2^{\log(k)^\alpha})$. What about LCCs?
- Best lower bound for 3-LDC: $n > k^2$
- Best lower bound for q -LDC: $n > k^{1+1/q-1}$
- Improve for LCCs !