UPPER BOUNDS

for

semi-RANDOM KSAT REFUTATIONS

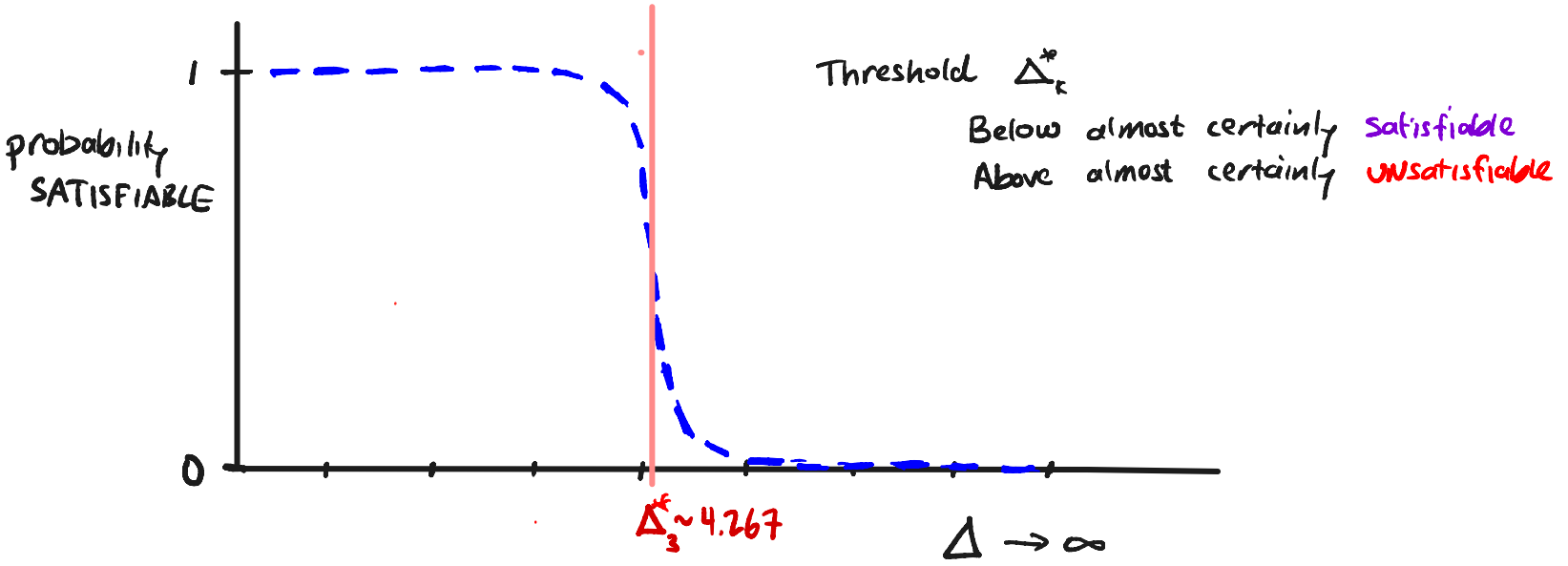
&

Lower Bounds for LDC's, LCC's

RANDOM K-CNFs

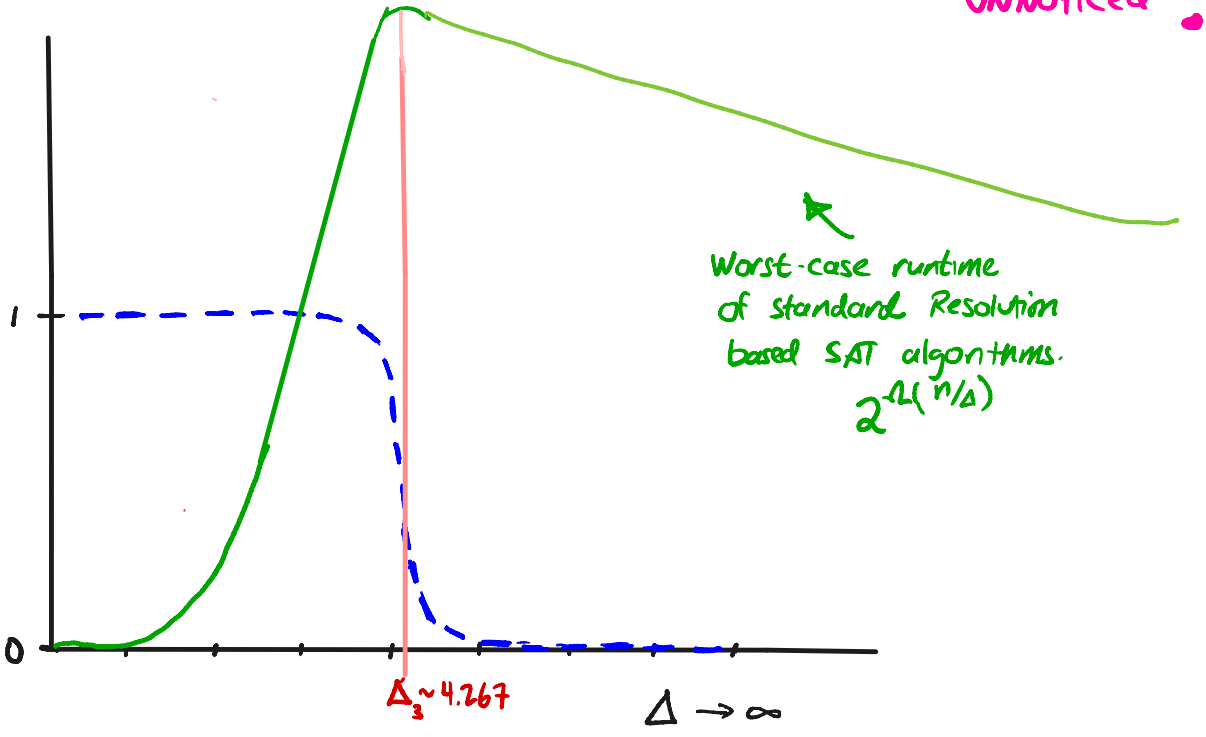
$f \sim \mathcal{F}(\Delta, n, k)$: pick $m = \Delta n$ clauses of width k

for $\Delta > 0$ suff large $f \sim \mathcal{F}(\Delta, n, k)$ UNSAT w.h.p.



Resolution-Based Algorithms for random KCNF

passing transition point Δ_k^* goes absolutely unnoticed !



OTHER RANDOM CNF FAMILIES

1. Random kXOR, random kCSP

2. $\text{Clique}_g(k)$, $g \sim \mathcal{G}(n,p)$ $\quad p \sim n^{-2/k-1}$

   $p = 1/2$ Clique of size $\log n$

3. $\text{Hard}_f(s)$ $\quad f \sim$ all boolean functions on n variables

   Says f computed by a size s circuit

MOTIVATION

1. Structural properties relate to our understanding

2. Natural distributions as benchmark for SAT algorithms

3. Lower bounds for particular proof systems (RES, SOS) give unconditional inapproximability for large family of algorithms

WHY IS IT SO HARD TO CERTIFY UNSAT OF RANDOM f?

Counting argument doesn't seem to work:

Circuit complexity:

$2^{poly(n)}$ circuits of poly size $\ll$ 2^{2^n} Boolean functions

Proof complexity

# of proofs of size s $\approx$ # UNSAT formulas

FEIGE'S HYPOTHESIS

Defn (Refutation algorithm)

Algorithm A is a refutation algorithm for random KSAT, $f \sim \mathcal{F}(\Delta, n, k)$:

A outputs YES with probability $> \frac{1}{2}$

A outputs NO if $\mathcal{Q}$ is satisfiable

Feige's Hypothesis: For $\Delta > 0$ sufficiently large, $f \sim \mathcal{F}(\Delta, n, k)$:

I. there is NO polytime refutation algorithm of f

II. NO proof system can efficiently refute f

The incredible usefulness of Feige's conjecture

Many problems are hard under Feige's conjecture:

- Approximating vertex cover
- Avg case MCSP
- PAC learning DNF

UPPER BOUNDS FOR RANDOM SAT

	Poly-size UB
Resolution	$m > n^2/\log n$ [Beame, Kautz, Pi, Saks] ($n^{k-1}/\log n$)
TC^0 Frege	$m \sim n^{1.4}$ [Feige, Kim, Ofek] [Müller, Tzameret]

LOWER BOUNDS FOR RANDOM SAT

	Poly-size UB	Expontial LB
Resolution	$m > n^2/\log n$ $k=3$ [Beame, Karp, P, Saks]	$m < n^{1.5}$ [Chvatal, Szemeredi] $k=O(1)$ [Beame, Karp, P, Saks] [Ben-Sasson, Wigderson]
Nullsatz		$m = O(n)$ [Grigoriev] $k=O(1)$
Poly Calculus		$m = O(n)$ $k=O(1)$ [Buss, Grigoriev, Impagliazzo, P]
SOS		$m = O(n)$ $k=O(1)$ [Gresoriev, Schoenebeck]
Cutting Planes		$k = \Theta(\log n)$ $m = \text{poly}(n)$ [Fleming, Pankratov, P, Robere/Hrubes, Pudlak]
TC° Frege	$m \sim n^{1.4}$ [Feige, Kim, Ofek] [Müller, Tzameret]	?

Refuting Semi-Random 3SAT

Random KSAT : Pick k-uniform hypergraph H over $\{x_1, \ldots x_n\}$ at random.

For each edge $C \in H$, randomly choose signs $b_1, \ldots b_k \in \{-1, 1\}$ of each literal.

whether variables in C occurs positively or negatively

Semi-random KSAT :

Fix arbitrary 3-hypergraph H over $\{x_1, \ldots x_n\}$, with m edges.

For each edge $C \in H$ randomly choose $b_1, b_2, b_3 \in \{-1, 1\}$

← hypergraph not random, only signs are random

Theorem whp there exists polysize Frege refutations for semirandom 3SAT instance, for $m \geq n^{1.4}$ clauses

Refuting Semi-Random 3SAT

Theorem whp there exists polysize Frege refutations for semirandom 3SAT instance, for $m \geq n^{1.4}$ clauses

[Feige, Kim, Ofek]
[Guruswami Kothari, Manohar '22]

Proof Plan:

I. Reduce weak refutation for semirandom 3SAT to (semi)-strong refutation for 3XOR via Feige XOR trick:

Theorem 2 Strong refutations for semi-random 3XOR with $m = n^{1.4}$ (show $val(f) < 1 - \frac{1}{n^2}$ whp) implies weak refutations of semi-random 3SAT, $m = n^{1.4}$

II. **Theorem 1**: $\exists$ strong refutations of semi-random 3XOR via Moore hypergraph bound.

Semi-strong Refutation for 3XOR

Theorem 1

Let $H = \{C_1, \ldots, C_m\}$ be arbitrary 3-uniform hypergraph over $[n]$

Let Ψ_H be semi-random 3XOR given by 3XOR constraints $(H, \vec{b}) = \{C_1 = b_1, C_2 = b_2, \ldots, C_m = b_m\}$, where $b_1 \ldots b_m \sim \{0,1\}$ are random.

Then for $m \geq 100\, n \left(\frac{n}{\ell}\right)^{1/2}$, whp over $b_1 \ldots b_m \sim \{0,1\}$:

$\mathrm{val}(\Psi_H) \triangleq$ max fraction of satisfied constraints of $\Psi_H \leq 1 - o\left(\frac{1}{\ell \log n}\right)$

- To refute Ksat via strong KXOR refutations, we will set $\ell \sim n^{1/5}$

Hypergraph Moore Bound

$k=2$ (ordinary graphs): Any graph with $\geq \frac{nd}{2}$ edges has a cycle of length $\sim 2\log_{d-1} n$

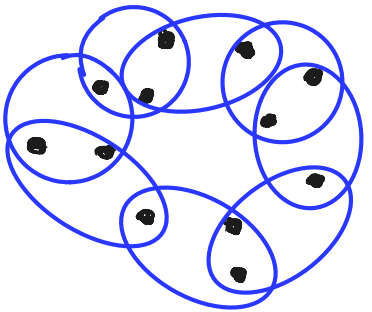
Hypergraph Moore Bound

$k=2$ (ordinary graphs): Any graph with $\geq \frac{nd}{2}$ edges has a cycle of length $\sim 2\log_{d-1} n$ [Alon-Hoory-Linial 2002]

Generalization to k-uniform hypergraphs:

A "cycle" is an even cover: set of k-edges $\mathcal{H}' \subseteq \mathcal{H}$ such that every vertex is contained in an even number of edges in E'

$k=3$



Hypergraph Moore Bound

Feige conjecture (2008):

Every k-uniform hypergraph H with $m \sim n\left(\frac{n}{\ell}\right)^{\left(\frac{k}{2}-1\right)}$ edges contains an even cover of length $\leq \ell \log_2 n$

↑

Proven up to polylog n factors — Guruswami - Kothari - Manohar '21; Hsieh - Kothari - Mohanty '22; H-K-M-Correlu-Sudakov '24

(We will sketch a simple proof time-permitting)

Semi-strong Refutation for semi-random 3XOR

Theorem 1 Let $H = \{C_1, \ldots, C_m\}$ be arbitrary 3-uniform hypergraph over $[n]$

Let Ψ_H be semi-random 3XOR given by 3XOR constraints $(H, \vec{b})$

Then for $m \geq 100 \underbrace{n\left(\frac{n}{\ell}\right)^{1/2}}_{m_0}$, whp over $b_1 \ldots b_m \sim \{0,1\}$:

$val(\Psi_H) \triangleq$ max fraction of satisfied constraints of $\Psi_H \leq 1 - O\left(\frac{1}{\ell \log n}\right)$

Proof : H satisfies conditions of even cover Theorem (Moore bound)

(1) Find $\ell \log n$ length even cover.

(2) Remove all hyperedges in cover, Let $H' = H$ - even cover

H' still has $100 m_0 - \ell \log n \geq m_0$ edges

(3) Repeatedly apply even cover theorem, partitioning $.99 m_0$ hyperedges of H into disjoint even covers, each of size $\leq \ell \log n$

(4) since each even cover is linearly independent, $\sim \frac{1}{2}$ of the even covers will be unsatisfiable (RHS of equations will sum to 1 mod 2)

$\therefore$ in total at least $\frac{1}{2}\left(\frac{.99 m_0}{\ell \log n}\right)$ constraints must be falsified

$\therefore$ whp $val(\Psi_H) \leq 1 - O\left(\frac{1}{\ell \log n}\right)$ ∎

Proof of Hypergraph Moore Bound ($k=4$)

Hypergraph Moore Bound:
Every k-uniform hypergraph H with $m \sim n\left(\frac{n}{\ell}\right)^{\left(\frac{k}{2}-1\right)}$ edges contains an even cover of length $\leq \ell \log_2 n$

Warmup: $\ell = 1$. Let's show every 4-uniform H with $\geq n^2$ edges contains a $\sim \log n$ length even cover

Pf Let g be a graph on $\binom{n}{2}$ vertices.

$(i,j) \sim (k,\ell)$ iff $\{i,j,k,\ell\} \in H$ and $i,j \neq k,\ell$

edges in g in 1-1 correspondence to hyperedges in H

cycles in g in 1-1 correspondence with an even cover in H

$\therefore$ follows by graph moore bound

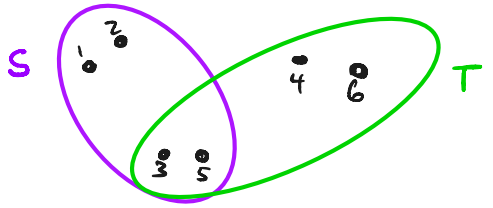
Proof of Hypergraph Moore Bound ($k=4$)

Let H be a 4-uniform hypergraph with $\geq n\left(\frac{1}{\ell}\right)^{\frac{k}{2}-1} \cdot \log n = \frac{n^2}{\ell} \log n$ edges.

Let $K_\ell(H)$ be the level-ℓ Kikuchi graph of H:

Vertices of $K_\ell(H)$: all $\binom{[n]}{\ell}$ ℓ-subsets of $[n]$

Edges of $K_\ell(H)$: (S,T) is an edge iff $S \oplus T \in H$



$(S,T) \in$ edge $(K_\ell(H))$

iff $\{1,2,4,6\} \in H$

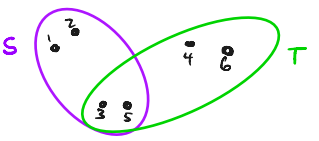
Proof of Hypergraph Moore Bound ($k=4$)

Let H be a 4-uniform hypergraph with $\geq n\left(\frac{n}{\ell}\right)^{\frac{k}{2}-1} \cdot \log n = \frac{n^2}{\ell}\log n$ edges.

Let $K_\ell(H)$ be the level-ℓ Kikuchi graph of H:

Vertices of $K_\ell(H)$: all $\binom{[n]}{\ell}$ ℓ-subsets of $[n]$

Edges of $K_\ell(H)$: (S,T) is an edge iff $S \oplus T \in H$



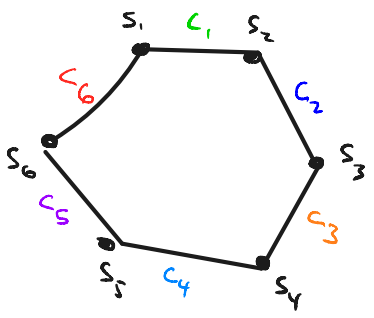
color edge (S,T) in $K_\ell(H)$ by $C = S \oplus T$

Claim

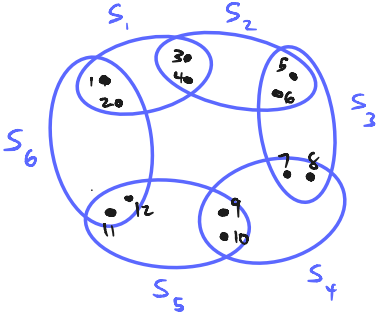
A closed walk in $K_\ell(H)$ where some color appears an odd # of times $\rightarrow$ even cover in H

(even cover = set of C_i's that occur an odd # of times in walk)

$k=4$
$\ell=4$



closed walk in $K_\ell(H)$



$S_1 \oplus S_2 = C_1 = \{1,2,5,6\}$

$S_2 \oplus S_3 = C_2 = \{3,4,7,8\}$

$S_3 \oplus S_4 = C_3 = \{5,6,9,10\}$

$S_4 \oplus S_5 = C_4 = \{7,8,11,12\}$

$S_5 \oplus S_6 = C_5 = \{9,10,1,2\}$

$S_6 \oplus S_1 = C_6 = \{11,12,1,2\}$

even cover in H

Proof of Hypergraph Moore Bound ($k=4$)

Thus it suffices to prove the following Lemma:

MAIN LEMMA: Let H be 4-uniform hypergraph with $\geq \frac{n^2}{\ell} \log n$ edges. Let $g = K_2(H)$ be colored Kikuchi graph for H. Then g has a closed walk of length $\leq \ell \log n$ where each color on walk occurs exactly once.

rainbow walk

Lemma Let g have nd edges. Then g contains a subgraph $g' \leq g$ with minimum degree $d' \geq d/4$ and at least $nd/2$ edges.

Proof of Hypergraph Moore Bound ($k=4$)

MAIN LEMMA: Let H be 4-uniform hypergraph with $\geq \frac{n^2}{\ell}\log n$ edges.
Let $g = K_\ell(H)$ be colored Kikuchi graph for H. Then g has a closed walk of length $\leq \ell \log n$ where each color on walk occurs exactly once.
rainbow walk

Pf (double count rainbow paths of length ℓ in g)

G has $N = \binom{n}{\ell}$ vertices

Edges: each $C \in H$ contributes $\binom{4}{2}\binom{n-4}{\ell-2}$ edges to G.

$\therefore$ G has $\geq \frac{n^2}{\ell}\binom{n}{\ell-2}\log n \geq 20N\log N$ edges, so avg degree $\sim 20\log N \sim 20\ell\log n$

Assume fsoc that G contains no short closed rainbow walks.

Let $G' \subseteq G$ be subgraph guaranteed by lemma, mindegree $d' \geq 5\log N$

Let $q = \log N \sim \ell \log n$

(i) # of length-q rainbow paths in G $\geq n \cdot d' \cdot (d'-1)\cdot(d'-2)\cdot\ldots\cdot(d'-(q-1)) \geq n(.9d')^q$

(ii) # length-q rainbow paths $\leq N^2 \cdot q! \approx 4^q \cdot q^q = (4q)^q$ ← if $\exists$ a closed rainbow walk then set of colors on every rainbow walk must use same set of q colors

$u \bullet \sim\sim \bullet v$

Contradiction since $4q = 4\log N < (.9)5\log N = .9d'$

Theorem 2 | Semi-strong KXOR Refutations $\longrightarrow$ Weak KSAT Refutations

Semirandom KSAT:

Fix arbitrary 3-hypergraph H over $\{x_1 \ldots x_n\}$, with m edges.

For each edge $C \in H$ randomly choose $b_1^C, b_2^C, b_3^C \in \{-1,1\}$

For each clause (C, b_1, b_2, b_3) its Fourier representation over $\mathbb{F}_2$ $(x_i \in \{-1,1\})$ is:

$$P(C,b_1,b_2,b_3) \stackrel{d}{=} \frac{7}{8} + \frac{1}{8}\left(b_1x_1 + b_2x_2 + b_3x_3 + b_1b_2x_1x_2 + b_1b_3x_1x_3 + b_2b_3x_2x_3 + b_1b_2b_3x_1x_2x_3\right)$$

Example: $C = \{x_1, x_2, x_3\}$ $b_1 = b_2 = b_3 = -1$ so clause is $(x_1 \vee x_2 \vee x_3)$

Fourier representation $= \frac{7}{8} + \frac{1}{8}\left(-x_1 - x_2 - x_3 - x_1x_2 - x_1x_3 - x_2x_3 - x_1x_2x_3\right)$

Defn

Let $\psi_H = \{(C_1, b^{C_1}), (C_2, b^{C_2}), \ldots (C_m, b^{C_m})\}$ be a semirandom 3SAT

Then $val(\psi) = \frac{1}{m} \cdot \max_{x \in \{-1,1\}^n} \sum_{i=1}^{m} P(C_i, b^{C_i})$ ← $val(\psi)$ is max fraction of satisfied clauses in ψ

Defn

Let $\psi_H = \{ (C_1, b^{c_1}), (C_2, b^{c_2}), \ldots (C_m, b^{c_m}) \}$ be a semirandom 3SAT

Then $val(\psi) = \max_{x \in \{-1,1\}^n} \underbrace{\frac{1}{m} \sum_{i=1}^{m} P(C_i, b^{c_i})}_{P}$ ← $val(\psi)$ is max fraction of satisfied clauses in ψ

write P as sum of 8 polynomials:

P_0 = all constant terms
P_1 = all linear terms
P_2 = all quadratic terms
P_3 = all degree 3 terms (3XORs)

$$val(\psi) \leq \underbrace{\max_x P_0}_{\frac{7}{8}} + \underbrace{\max_x P_1}_{\frac{1}{8} \cdot O(\sqrt{n/m}\sqrt{\log n})} + \underbrace{\max_x P_2}_{\frac{1}{8} \cdot O(\sqrt{n/m}\sqrt{\log n})} + \underbrace{\max_x P_3}_{\frac{1}{8}(1 - O(\frac{1}{\ell \log n}))}$$

we'll show:

$$val(\psi) \le \underbrace{\max_x P_0}_{\frac{7}{8}} + \underbrace{\max_x P_1}_{\frac{1}{8}\cdot O\left(\sqrt{n/m}\sqrt{\log n}\right)} + \underbrace{\max_x P_2}_{\frac{1}{8}\cdot O\left(\sqrt{n/m}\sqrt{\log n}\right)} + \underbrace{\max_x P_3}_{\frac{1}{8}\left(1-O\left(\frac{1}{\ell \log n}\right)\right)}$$

we'll show:

Assuming these upper bounds, and $m = O(n\sqrt{\frac{n}{\ell}}\log n)$

$$val(\psi) \le \frac{7}{8} + \left(\frac{\ell}{n}\right)^{1/4} + \frac{1}{8}\left(1-O\left(\frac{1}{\ell \log n}\right)\right) = 1 + \left(\frac{\ell}{n}\right)^{1/4} - O\left(\frac{1}{\ell \log n}\right)$$

this is < 1 if $\left(\frac{\ell}{n}\right)^{1/4} \ll \frac{1}{\ell \log n}$.

This happens if $\ell^{5/4} \ll n^{1/4}$, so setting $\ell \le n^{1/5}$ achieves this.

Choosing $\ell = n^{1/5}$ gives $m = n^{1.4}$.

So it is left to prove the claimed upper bounds.

Degree 3 terms (3XOR2 part)

upper bound on $\mathrm{val}(P_3)$ follows by Theorem 1 !

UPPER bounds for linear part (P_1) and quadratic part (P_2) is easier.
We sketch proofs of these next.

Linear terms

Say x_i occurs in n_i many clauses.
Because signs are random, the coefficient in front of x_i has expectation $\sim \frac{\sqrt{n_i}}{m}$

$$\therefore P_1 = \frac{1}{m} \sum_i \sqrt{n_i}\, x_i$$

$$\therefore \max_{x \in \{-1,1\}^n} P_i \le \frac{n}{m} \frac{1}{n} \sum_i \sqrt{n_i} \le \frac{n}{m} \sqrt{\frac{1}{n} \sum_i n_i} = \frac{n}{m} \sqrt{\frac{m}{n}} = O\left(\sqrt{\frac{n}{m}}\right)$$

(Cauchy-Swartz)

Quadratic terms

Key idea: we can write quadratic part as $\frac{1}{m} x^T \left(\sum_{c \in H} a_c A_c \right) x$

where A_c = $n \times n$ matrix with 1 in (i,j) iff $x_i, x_j \in C$

a_c = coefficient of $x_i x_j$ in Fourier expansion of clause (C, b^C)

Note nonzero entries of $\sum_c a_c A_c$ are determined by H, but sign (± 1) is random.

$\therefore \frac{1}{m} x^T \left(\sum_c a_c A_c \right) x \leq \frac{n}{m} \left\| \sum_c a_c A_c \right\|_2$ since $\|x\|_2 = \sqrt{n}$

By Matrix Khintchine (Matrix Chernoff bound):

$$\left\| \sum_c a_c A_c \right\|_2 \leq \left(\sum_c A_c^2 \right)^{1/2} \sqrt{\log m} \leq \sqrt{\frac{m}{n}} \sqrt{\log n}$$

assuming H is random. more complicated arg if H not random.

$\therefore \max_x p_2 \leq \frac{n}{m} \left\| \sum_c a_c A_c \right\|_2 \leq \frac{n}{m} \sqrt{\frac{m}{n}} \sqrt{\log n} = \sqrt{\frac{n}{m}} \sqrt{\log n}$

Remarks

(1) The whole proof can be formalized in poly sized Frege pf.

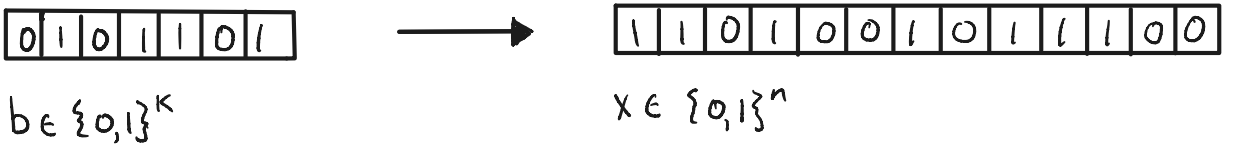
The hard part (Theorem 1) actually formalized in much weaker system – poly-size Poly calculus (PC) refutation over $\mathbb{F}_2$

(2) No improvements to $m \geq n^{1.4}$ given in original FKO paper.

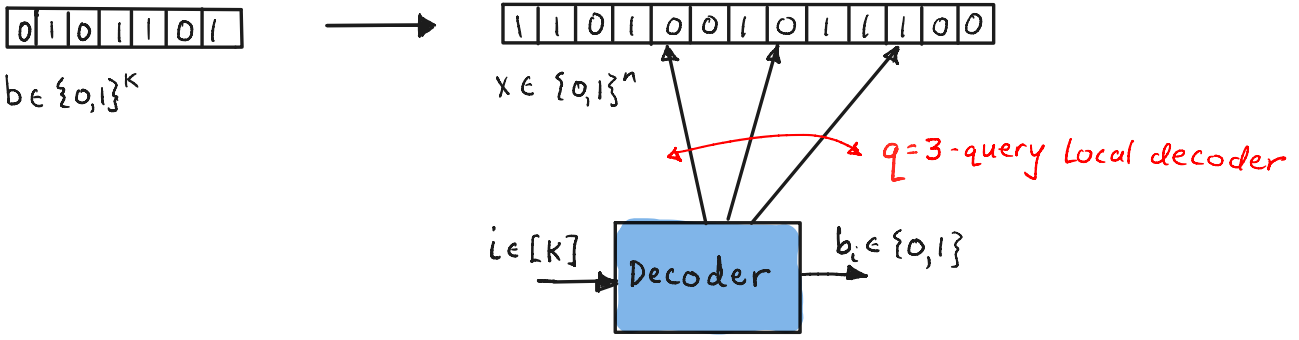
(3) Strong LBs for Resolution refutations:

for $m \leq n^{5/4 - \varepsilon}$ 3clauses, Resolution refutations require exponential size

Locally Decodable Codes



Locally Decodable Codes

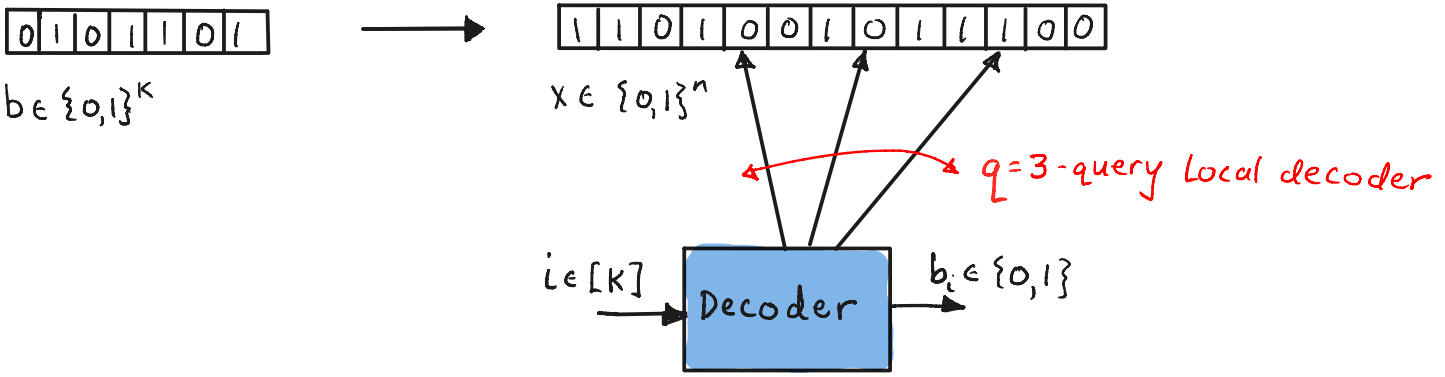


(q, ε, δ)-LDC: given received word x with $\leq \delta$ fraction of errors, for any position i, $\text{Decoder}(i, x) = x_i$ with probability $\geq 1 - \varepsilon$

Applications: PCP's, Private Information Retrieval, secret sharing, worst-to-avg case reductions, Distributed computation, ...

Open: Does there exist $q = O(1)$ LDC with $n = \text{poly}(K)$?

Locally Decodable Codes



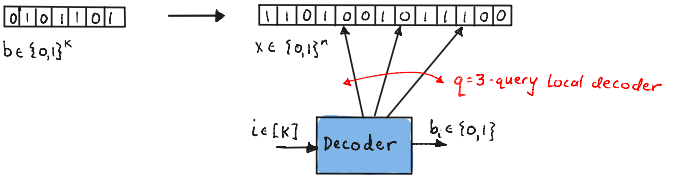
Best construction $K \longrightarrow 2^{K^\varepsilon}$ "Matching Vector Codes" [Yek'08, Efr09, DGY'11]

New Lower Bounds

$n = \Omega(K^3)$ for 3-query LDC's [AGKM23, Yankovitz 24]

$n = 2^{\Omega(K)}$ for linear 3-query LCC's (locally correctable) [Kothari, Manohar '23] [Alrabiah, Guruswami, '24]

SEMI-RANDOM XORs & LDC LOWER BOUNDS

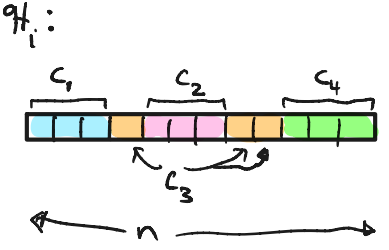


* Breakthrough Lower Bounds: formalized as system of semi-random XOR constraints: $\mathcal{F} = \{F_\beta \in \{0,1\}^K : F_\beta\}$

Normal Form [Yek'08]

$\exists$ 3-uniform hypergraph matchings $\mathcal{H}_1, \ldots, \mathcal{H}_K$, each $\mathcal{H}_i$ over $\{1, \ldots, n\}$

Decoding: on $i \in [K]$ pick random $C \in \mathcal{H}_i$, output $\sum_{v \in C} x_v \bmod 2$



System of XORs: $\forall \beta \in \{0,1\}^K$: $F_\beta = \{\forall i \in [K], C \in \mathcal{H}_i : \sum_{v \in C} x_v = \beta_i\}$

Lemma: F_β highly UNSAT for random β $\Rightarrow$ LB $\Omega(n)$ for LDC's

Proof based on ideas in semirandom CSP refutations