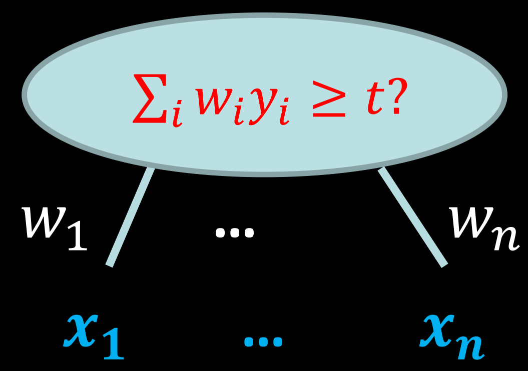
Some Depth Two (and Three) Threshold Circuit Lower Bounds

Ryan Williams **Stanford**
Joint work with Daniel Kane (UCSD)

Introduction

Def. $f_n\colon \{0,1\}^n \to \{0,1\}$ is a linear threshold function (LTF) if there are $w_1, \dots w_n, t \in \mathbb{R}$ such that

$$\forall\ (x_1, \dots, x_n) \in \{0,1\}^n,\ \ f(x_1, \dots, x_n) = 1 \ \text{ iff }\ \sum_i w_i x_i \geq t.$$



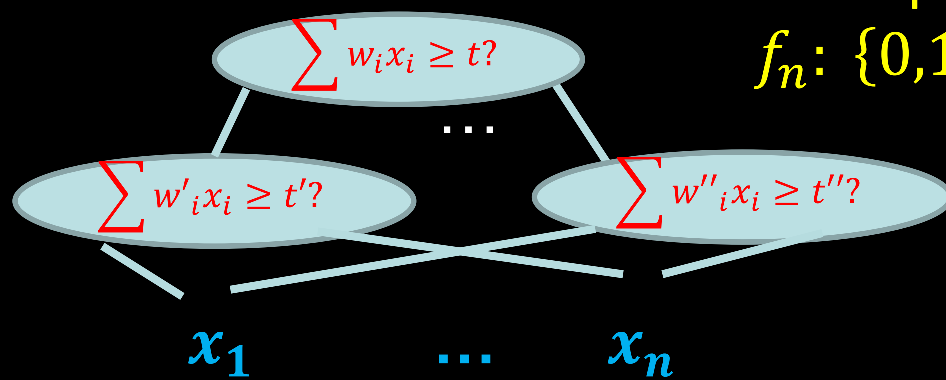
[McCulloch-Pitts ’40s, Minsky-Papert '60s]

Depth-Two LTF Circuits

"Multilayer perceptron with one hidden layer"

This talk: "LTF ∘ LTF"

Computes some $f_n \colon \{0,1\}^n \to \{0,1\}$



[MP'60s] Every $f_n \colon \{0,1\}^n \to \{0,1\}$ is computable by a LTF ∘ LTF

But the circuit could have **$2^{\Omega(n)}$ gates/wires**…!

Question: Which functions $f = \{f_n\}$ have a LTF ∘ LTF circuit family with $n^{O(1)}$ gates?

Which do not?

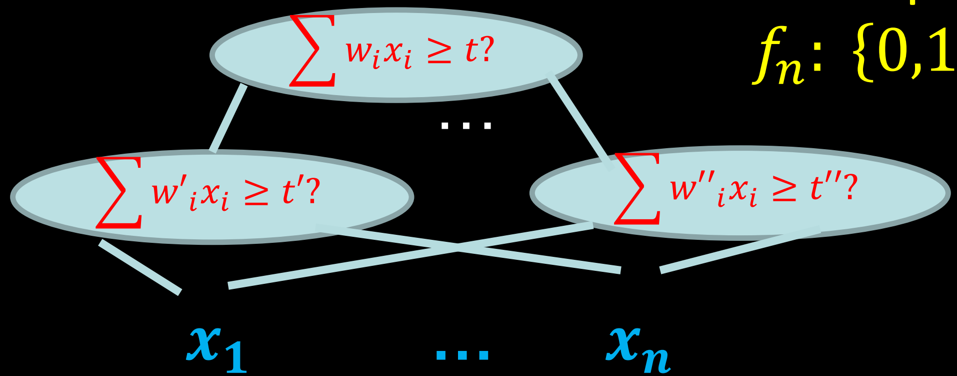
What about $O(n)$ gates?

Depth-Two LTF Circuits

"Multilayer perceptron with one hidden layer"

This talk: "LTF ∘ LTF"

Computes some $f_n\colon \{0,1\}^n \to \{0,1\}$



- [IPS'93] PARITY of n bits needs $\Omega\left(n^{1.5}\right)$ wires
- [ROS'93] Inner Prod Mod 2 on n-bit vectors needs $\Omega(n)$ gates
- If the weights in one layer are restricted to the set {-1,0,1}, get **exp(n)** lower bounds [HMPST'93], [Nisan'94], [FKLMSS'01]
  "TC0 depth-2 lower bounds"

OPEN: $E^{NP} \subset$ (LTF ∘ LTF of O(n) gates and O($n^{1.6}$) wires)?
TC0 depth-3 circuits of O(n) gates and O($n^{1.6}$) wires?

This Work: Some Lower Bounds

Theorem 1 There's a function A : **{0,1}*** → **{0,1}** such that

1. A ∈ **P** (in fact, A is in uniform **TC0 depth-3** of $O(n)$ gates, and A has LTF ∘ LTF circuits with $O\left(n^3\right)$ gates)
2. For all n, A_n does not have **LTF ∘ LTF circuits** with $n^{1.5}/polylog(n)$ **gates**, nor with $n^{2.5}/polylog(n)$ wires (even for a ½+o(1) fraction of inputs!)

Theorem 2 There's a function B : **{0,1}*** → **{0,1}** such that

1. B ∈ **P**
2. For all n, B_n does not have **TC0 depth-3 circuits** of $n^{1.5}/polylog(n)$ gates, nor with $n^{2.5}/polylog(n)$ wires

Remainder of the Talk: Size ≡ Gates

Outline

- Intuition Behind Theorem 1
- LTF-LTF Lower Bound for *Random* Functions
- Random Restriction Lemma for LTFs
- Theorem 2 (Briefly)
- Conclusion

Andreev's Function [A'87]

The $\boldsymbol{k}$**-bit multiplexer function** is defined as:

$$M_{2^k}(x_1, \dots, x_{2^k}, a_1, \dots, a_k) := x_{bin(a_1 \cdots a_k)}$$

Let k be a power of two. **Andreev's function** on $\boldsymbol{n = 2^{k+1}}$ bits:

$$A_n\left(x_1, \dots, x_{2^k}, a_{1,1}, \dots, a_{1,2^k/k}, \dots, a_{k,1}, \dots, a_{k,2^k/k}\right)$$
$$:= M_{2^k}(x_1, \dots, x_{2^k}, \bigoplus_j a_{1,j}, \dots, \bigoplus_j a_{k,j})$$

Theorem [A'87, H'98]

A_n **requires DeMorgan formulas of size** $\Omega\left(\frac{n^3}{polylogn}\right)$

Our proof of Theorem 1 will mimic aspects of this theorem.
So let's look at the intuition for it…

n^3 Formula Lower Bound for Andreev

$$M_{2^k}(x_1, \dots, x_{2^k}, a_1, \dots, a_k) := x_{bin(a_1 \cdots a_k)}$$

$$A_n\left(x_1, \dots, x_{2^k}, a_{1,1}, \dots, a_{1,2^k/k}, \dots, a_{k,1}, \dots, a_{k,2^k/k}\right) := M_{2^k}\left(x_1, \dots, x_{2^k}, \bigoplus_j a_{1,j}, \dots, \bigoplus_j a_{k,j}\right)$$

Define a random process R on Boolean functions f with n inputs:

R(f): Randomly choose all but 100 log n inputs to f.
Set all other inputs uniformly at random.
Output the (100 log n)-input function g.

n^3 Formula Lower Bound for Andreev

$$M_{2^k}(x_1, \ldots, x_{2^k}, a_1, \ldots, a_k) := x_{bin(a_1 \cdots a_k)}$$

$$A_n\left(x_1, \ldots, x_{2^k}, a_{1,1}, \ldots, a_{1,2^k/k}, \ldots, a_{k,1}, \ldots, a_{k,2^k/k}\right) := M_{2^k}(x_1, \ldots, x_{2^k}, \bigoplus_j a_{1,j}, \ldots, \bigoplus_j a_{k,j})$$

R(f): Randomly choose all but 100 log n inputs to f.
Set all other inputs uniformly at random.
Output the (100 log n)-input function g.

IDEA 1: On a DeMorgan formula F of size s, process $R(F)$ outputs function g of expected size < $s \cdot \frac{polylog\ n}{n^2} + \sqrt{s} \cdot \frac{1}{n}$ **[H'98]**

IDEA 2: $R(A_n) = g$ needs $\geq \frac{n}{polylog\ n}$ size formulas, whp **[A'87]**

IDEA 2a: A *random* function $f: \{0,1\}^k \to \{0,1\}$ needs $\geq \frac{2^k}{\log k}$ size, whp

IDEA 2b: As $R(A_n)$ assigns $x_1, \ldots, x_n$ *random* 0/1 values, the output g needs $\geq \frac{n}{\log \log n}$ size, whp. **(whp for all i, some $a_{i,j}$ is unset)**

n^3 Formula Lower Bound for Andreev

$$M_{2^k}(x_1,\ldots,x_{2^k},a_1,\ldots,a_k) := x_{bin(a_1\cdots a_k)}$$

$$A_n\left(x_1,\ldots,x_{2^k},a_{1,1},\ldots,a_{1,2^k/k},\ldots,a_{k,1},\ldots,a_{k,2^k/k}\right) := M_{2^k}(x_1,\ldots,x_{2^k},\bigoplus_j a_{1,j},\ldots,\bigoplus_j a_{k,j})$$

R(f): Randomly choose all but 100 log n inputs to f.
Set all other inputs uniformly at random.
Output the (100 log n)-input function g.

IDEA 1: On a DeMorgan formula F of size s, process $R(F)$ outputs function g of expected size < $s \cdot \frac{polylog\, n}{n^2} + \sqrt{s} \cdot \frac{1}{n}$ **[H'98]**

IDEA 2: $R(A_n) = g$ needs $\geq \frac{n}{polylog\, n}$ size formulas, whp **[A'87]**

Combining 1 and 2: If F computes A_n, then $s \geq \frac{n^3}{polylog\, n}$

Theorem **A_n requires DeMorgan formulas of size $\Omega\left(\frac{n^3}{polylog\, n}\right)$**

n^3 Formula Lower Bound for Andreev

$$M_{2^k}(x_1, \ldots, x_{2^k}, a_1, \ldots, a_k) := x_{bin(a_1 \cdots a_k)}$$

$$A_n\left(x_1, \ldots, x_{2^k}, a_{1,1}, \ldots, a_{1,2^k/k}, \ldots, a_{k,1}, \ldots, a_{k,2^k/k}\right) := M_{2^k}(x_1, \ldots, x_{2^k}, \oplus_j\, a_{1,j}, \ldots, \oplus_j\, a_{k,j})$$

R(f): Randomly choose all but 100 log n inputs to f.
Set all other inputs uniformly at random.
Output the (100 log n)-input function g.

IDEA 1: On a DeMorgan formula F of size s, process $R(F)$ outputs function g of expected size < $s \cdot \frac{polylog\, n}{n^2} + \sqrt{s} \cdot \frac{1}{n}$ **[H'98]**

IDEA 2: $R(A_n) = g$ needs $\geq \frac{n}{polylog\, n}$ size formulas, whp **[A'87]**

Analogues of Ideas 1 and 2 for depth-2 LTF circuits?

LTF-LTF Lower Bound for Andreev?

$M_{2^k}(x_1, \dots, x_{2^k}, a_1, \dots, a_k) := x_{bin(a_1 \cdots a_k)}$

$A_n\left(x_1, \dots, x_{2^k}, a_{1,1}, \dots, a_{1,2^k/k}, \dots, a_{k,1}, \dots, a_{k,2^k/k}\right) := M_{2^k}(x_1, \dots, x_{2^k}, \bigoplus_j a_{1,j}, \dots, \bigoplus_j a_{k,j})$

R(f): Randomly choose all but 100 log n inputs to f.
Set all other inputs uniformly at random.
Output the (100 log n)-input function g.

Lemma 1: On a LTF-LTF circuit C of s size (gates), $R(C)$ outputs a function g of expected LTF-LTF size < $s \cdot \frac{O(\log n)}{n^{1/2}}$

Lemma 2: $R(A_n)$ = g needs $\geq \frac{n}{polylog\ n}$ size LTF-LTF circuits, whp

Combining 1 and 2: If C computes A_n, then $\boldsymbol{s} \geq \frac{n^{1.5}}{polylog\ n}$

Theorem 1 A_n requires LTF-LTF circuits of size $\Omega\left(n^{1.5-o(1)}\right)$

Outline

- Intuition Behind Theorem 1
- LTF-LTF Lower Bound for *Random* Functions
- Random Restriction Lemma for LTFs
- Theorem 2 (Briefly)
- Conclusion

LTF-LTF Lower Bound for Random f's

$M_{2^k}(x_1, \ldots, x_{2^k}, a_1, \ldots, a_k)$
$:= x_{bin(a_1 \cdots a_k)}$

$A_n\left(x_1, \ldots, x_{2^k}, a_{1,1}, \ldots, a_{1,2^k/k}, \ldots, a_{k,1}, \ldots, a_{k,2^k/k}\right)$
$:= M_{2^k}(x_1, \ldots, x_{2^k}, \bigoplus_j a_{1,j}, \ldots, \bigoplus_j a_{k,j})$

R(f): Randomly choose all but 100 log n inputs to f.
Set all other inputs uniformly at random.
Output the (100 log n)-input function g.

Reminder of Lemma 2:

$R(A_n) = g$ needs $\geq \frac{n}{polylog\ n}$ size LTF-LTF circuits, whp

Lemma 2 follows from:

Theorem: *Random* $f: \{0,1\}^k \rightarrow \{0,1\}$ need $\geq \frac{2^k}{poly(k)}$ gates to be computed by LTF-LTF circuits, whp

Want to show an exponential lower bound for random functions

LTF-LTF Lower Bound for Random f's

Theorem: *Random* $f: \{0,1\}^k \rightarrow \{0,1\}$ need $> \frac{2^k}{k^3}$ gates to be computed by LTF-LTF circuits, whp

Key Claim: The number of distinct Boolean functions $f: \{0,1\}^n \rightarrow \{0,1\}$ computable by LTF-LTF circuits of s gates is $\mathbf{2^{O(n^2 s)}}$

Proof of Theorem (Assuming Claim): The number of functions $f: \{0,1\}^n \rightarrow \{0,1\}$ computable by LTF-LTF circuits of $s = \frac{2^n}{n^3}$ gates is only $\mathbf{2^{O\left(n^2 \cdot \left(\frac{2^n}{n^3}\right)\right)}} \leq \mathbf{2^{\frac{2^n}{n}}}$. But there are 2^{2^n} Boolean functions.

So, $\Pr_f[f \text{ has size s LTF}-\text{LTF circuits}] = o(1)$. **QED**

LTF-LTF Lower Bound for Random f's

Key Claim: The number of distinct Boolean functions $f: \{0,1\}^n \to \{0,1\}$ computable by LTF-LTF circuits of s gates is $\mathbf{2^{O(n^2 s)}}$

First consider the case s = 1. We'll use:

Chow's Theorem [FOCS'61] Every LTF f is uniquely determined by the Fourier coefficients $\hat{f}(\emptyset), \hat{f}(\{1\}), \ldots, \hat{f}(\{n\})$

How do we use this? Well, what *is* a Fourier coefficient?

For $f: \{-1,1\}^n \to \{-1,1\}$ and $S \subseteq [n]$,

$\hat{f}(S) = E_x[f(x) \cdot PARITY_S(x)]$

So $\hat{f}(S) = j/2^n$, where $j \in \{-2^n, \ldots, -1, 0, 1, \ldots, 2^n\}$

Thus there are $\left(2^{n+1}+1\right)^{n+1} \leq 2^{O(n^2)}$ choices for the Fourier coefficients, so $2^{O(n^2)}$ LTFs on n variables

LTF-LTF Lower Bound for Random f's

Key Claim: The number of distinct Boolean functions $f: \{0,1\}^n \to \{0,1\}$ computable by LTF-LTF circuits of s gates is $\mathbf{2^{O(n^2 s)}}$

Theorem (see also [ROS'94]) Fix $f_1, \ldots, f_s$: $\{0,1\}^n \to \{0,1\}$. There are $\mathbf{2^{O(ns)}}$ functions of the form $h(x) = g(f_1(x), \ldots, f_s(x))$, where g is an LTF on s bits.

Proof Idea. Show that every such h is uniquely determined by:

1. The ***number*** of distinct vectors $y \in \{0,1\}^s$ such that $y = \left(f_1(x), \ldots, f_s(x)\right)$ and $g(y) = 1$
2. The ***component-wise sum*** of all such y's

Total number of possibilities is $\mathbf{2^{O(ns)}}$

➔ **Total number of LTF-LTF functions is $\left(2^{O(n^2)}\right)^s \cdot 2^{O(n\,s)}$**

Outline

- Intuition Behind Theorem 1
- LTF-LTF Lower Bound for *Random* Functions
- Random Restriction Lemma for LTFs
- Theorem 2 (Briefly)
- Conclusion

Random Restrictions to LTFs

R(f): Randomly choose all but 100 log n inputs to f.
Set all other inputs uniformly at random.
Output the (100 log n)-input function g.

Reminder of Lemma 1: On a LTF-LTF circuit C of s size (gates), $R(C)$ outputs a function g of expected LTF-LTF size < $s \cdot \frac{O(\log n)}{n^{1/2}}$

(Weak) Random Restriction Lemma: Let f be an LTF on n vars. Let k << n. Randomly choose all but k inputs to f and assign all other inputs randomly, obtaining g on k vars. Then

$$\mathbf{Pr}[g \text{ is a constant function}] \geq 1 - \frac{O(k)}{\sqrt{n}}$$

Proof of Lemma 1 (Assuming RR Lemma):

Set k = 100 log n. After applying R to circuit C, the expected number of *non-constant* gates on the bottom layer is $s \cdot \frac{O(\log n)}{n^{1/2}}$

(Weak) Random Restriction Lemma: Let f be an LTF on n vars. Let k << n. Randomly choose all but k inputs to f and assign all other inputs randomly, obtaining g on k vars. Then

$$\textbf{Pr}[g \text{ is a constant function}] \geq 1 - \frac{O(k)}{\sqrt{n}}$$

Remark: The lemma is already tight for f = MAJORITY.

Randomly assign 0/1 to all but one input of f.

What's the probability the remaining bit influences the output?

$$\Pr[\text{½ of the other bits are set 0, ½ are set to 1}] \sim \frac{1}{\sqrt{n}}$$

Intuition: When does an LTF become a constant function?

When its threshold value becomes "too high" or "too low" after a partial assignment to the variables

Consider LTF defined by linear form $L(x)$ and threshold value t.

Let $B \subseteq [n]$ be the index set of the k *unassigned* vars.

Randomly assign $x_i \coloneqq v_i$ for all $i \notin B$. Let $L'(v) = \sum_{j \notin B} a_j v_j$.

If $L'(v) < t - \sum_{i \in B:\, a_i > 0} |a_i|$ then the remaining LTF $\equiv 0$.

If $L'(v) > t + \sum_{i \in B:\, a_i < 0} |a_i|$ then the remaining LTF $\equiv 1$.

[LO'43, E'45] Let $L(x_1, \dots, x_n) = \sum_j a_j x_j$. Let I be an (open) interval of $\mathbb{R}$. Suppose there are k integers j such that $|a_j| \geq |I|$. Then $\Pr_{x \in \{0,1\}^n}[L(x) \in I] \leq \frac{O(1)}{\sqrt{k}}$

Proof [Erdős'45]: Note WLOG, all a_i are positive.
Fix 0/1 values for all x_j with $a_j < |I|$. For the remaining variables $(x_{i_1}, \dots, x_{i_k}) \in \{0,1\}^k$ with $a_{i_j} \geq |I|$, define

$$S_{x_{i_1},\dots,x_{i_k}} := \{ j \mid x_{i_j} = 1\} \subseteq [k].$$

Note if $L(x)$ and $L(y)$ are in I, then $|L(x) - L(y)| < |I|$.
This implies the corresponding $S_{x_{i_1},\dots,x_{i_k}}$ and $S_{y_{i_1},\dots,y_{i_k}}$ are ***incomparable.***
Therefore $S' = \left\{S_{x_{i_1},\dots,x_{i_k}} \middle| L(x) \in I\right\}$ is an anti-chain, so $|S'| \leq \binom{k}{k/2}$.
There are 2^k total assignments to the $(x_{i_1}, \dots, x_{i_k})$.
But this holds for *all 0/1* choices of x_j with $a_j < |I|$, so

$$\Pr_{x \in \{0,1\}^n}[L(x) \in I] \leq \frac{\binom{k}{k/2}}{2^k} \leq \frac{O(1)}{\sqrt{k}}$$

(Weak) Random Restriction Lemma: Let f be an LTF on n vars. Let k << n. Randomly choose all but k inputs to f and assign all other inputs randomly, obtaining g on k vars. Then

$$\mathbf{Pr}[g \text{ is a constant function}] \geq 1 - \frac{O(k)}{\sqrt{n}}$$

A simple case: Suppose all $a_i \in \{-1, 1\}$.

Consider LTF defined by linear form $L(x)$ and threshold value t.

Let $B \subseteq [n]$ be *any* index set of k *unassigned* vars.

Randomly assign $x_i := v_i$ for all $i \notin B$. Let $L'(v) = \sum_{j \notin B} a_j v_j$.

The remaining LTF on the vars in B is *not* a constant function

$$\Leftrightarrow \quad L'(v) \in (t - P, t + Q),$$

where P = (# $i \in B$: $a_i = 1$) and Q = (# $i \in B$: $a_i = -1$).

Want to upper bound: $Pr_v[\sum_{j \notin B} a_j v_j \in (t - P, t + Q)]$

Idea: Divide $(t - P, t + Q)$ into k intervals $I_1, \ldots, I_k$ of length 1.

For every i, we have $Pr_v[L'(v) \in I_i] \leq \frac{O(1)}{\sqrt{n-k}}$ by the L-O Lemma.

Then Pr[…] $\leq \sum_i Pr_v[L'(v) \in I_i] \leq \frac{O(k)}{\sqrt{n-k}}$ by the union bound

Outline

- Intuition Behind Theorem 1
- LTF-LTF Lower Bound for *Random* Functions
- Random Restriction Lemma for LTFs
- Theorem 2 (Briefly)
- Conclusion

Theorem 2 There's a function $B : \{0,1\}^* \to \{0,1\}$ such that

1. $B \in$ P
2. For all n, B_n does not have **TC0 depth-3 circuits** of $n^{1.5}/polylog(n)$ **gates** nor with $n^{2.5}/polylog(n)$ wires

Observation: B_n **CANNOT BE** A_n

A_n **has TC0 depth-3 circuits of $O(n)$ gates!**

Let D be an $O\left(\frac{n}{\epsilon^2}\right) \times n$ matrix, whose rows are n-bit strings in an $\varepsilon = \frac{1}{n^7}$-biased set. Think of $D: \mathrm{F}_2^n \to \mathrm{F}_2^{O\left(\frac{n}{\epsilon^2}\right)}$ as a linear code of distance ½-$\Omega(\epsilon)$.

Define:

$$B_n(x_1, \dots, x_n, a_{1,1}, \dots, a_{1,n/(15 \log n)}, \dots, a_{k,1}, \dots, a_{k,n/(15 \log n)}) := M_{n^{15}}(D(x_1, \dots, x_n), \oplus_j a_{1,j}, \dots, \oplus_j a_{k,j})$$

Theorem 2 There's a function $B : \{0,1\}^* \to \{0,1\}$ such that

1. $B \in$ P
2. For all n, B_n does not have **TC0 depth-3 circuits** of $n^{1.5}/polylog(n)$ **gates** nor with $n^{2.5}/polylog(n)$ wires

Let D be an $O\left(\frac{n}{\epsilon^2}\right) \times n$ matrix, whose rows are n-bit strings in an $\varepsilon = \frac{1}{n^7}$-biased set. Think of $D: \mathrm{F}_2^n \to \mathrm{F}_2^{O\left(\frac{n}{\epsilon^2}\right)}$ as a linear code of distance ½-$\Omega(\epsilon)$.

$$B_n\left(x_1, \dots, x_n, a_{1,1}, \dots, a_{1,n/(15\ log\ n)}, \dots, a_{k,1}, \dots, a_{k,n/(15\ log\ n)}\right)$$
$$:= \ M_{n^{15}}\left(D(x_1, \dots, x_n), \bigoplus_j a_{1,j}, \dots, \bigoplus_j a_{k,j}\right)$$

KEY IDEA: For *random* x, think of the output $y = D(x)$ as a (15 log n)-bit Boolean function f_y

Whp, every LTF-LTF circuit with 15 log n inputs and o(n/polylog n) gates *disagrees* with f_y on at least a $\left(\frac{1}{2} - \frac{1}{n^{2.6}}\right)$-fraction of inputs!

Therefore, after a random restriction to all but (15 log n) inputs, no Majority of $n^{2.5}$ LTF-LTFs of o(n/polylog n) size can compute B_n

Conclusion

- **Tight upper and lower bounds?**

  Currently $O(n^3)$ LTF-LTF circuits for Andreev,

  but have only $\Omega(n^{1.5})$ lower bound. Which is the truth?

- **LTF-LTFs of poly(n) size for Inner Product Mod 2?**

  Negative answer would separate TC0 depth-3 from LTF-LTF

- **SAT algorithms for LTF-LTF circuits?**

  [Chen-Santhanam-Srinivasan'16]

  2^{n-n^δ} time for SAT of LTF-LTF with $O(n^{1+\epsilon})$ wires

  [Alman-Chan-W'??] 2^{n-n^δ} time for ACC-LTF-LTF w/ $n^{2-\epsilon}$ gates

- **Lower bounds for PTF-PTF circuits? Nothing superlinear (yet)**
- **Fast evaluation of LTF-LTF → Circuit LBs??**

  [W'14] $2^n\, poly(n)$ time algorithm for evaluating an

  LTF-LTF circuit of $2^{\delta n}$ gates on all possible inputs

Thank you!