Super-Linear Gate and Super-Quadratic Wire Lower Bounds for Depth-Two and Depth-Three Threshold Circuits

Daniel M. Kane
University of California, San Diego

Ryan Williams[*]
Stanford University

Abstract

In order to formally understand the power of neural computing, we first need to crack the frontier of threshold circuits with two and three layers, a regime that has been surprisingly intractable to analyze.

We prove the first super-linear gate lower bounds and the first super-quadratic wire lower bounds for depth-two linear threshold circuits with arbitrary weights, and depth-three majority circuits computing an explicit function.

- We prove that for all $\varepsilon \gg \sqrt{\log(n)/n}$, the linear-time computable Andreev's function cannot be computed on a $(1/2 + \varepsilon)$-fraction of n-bit inputs by depth-two linear threshold circuits of $o(\varepsilon^3 n^{3/2}/\log^3 n)$ gates, nor can it be computed with $o(\varepsilon^3 n^{5/2}/\log^{7/2} n)$ wires.

  This establishes an average-case "size hierarchy" for threshold circuits, as Andreev's function is computable by uniform depth-two circuits of $o(n^3)$ linear threshold gates, and by uniform depth-three circuits of $O(n)$ majority gates.

- We present a new function in P based on small-biased sets, which we prove cannot be computed by a majority vote of depth-two linear threshold circuits with $o(n^{3/2}/\log^3 n)$ gates, nor with $o(n^{5/2}/\log^{7/2} n)$ wires.

- We give tight average-case (gate and wire) complexity results for computing PARITY with depth-two threshold circuits; the answer turns out to be the same as for depth-two majority circuits.

The key is a new random restriction lemma for linear threshold functions. Our main analytical tool is the Littlewood-Offord Lemma from additive combinatorics.

1 Introduction

A function $f : \{0,1\}^n \to \{0,1\}$ is a *linear threshold function* (LTF) if there are *weights* $w_1, \ldots, w_n, t \in \mathbb{R}$ such that for all $(a_1, \ldots, a_n) \in \{0,1\}^n$, $f(a_1, \ldots, a_n) = 1$ if and only if $\sum_i w_i a_i \geq t$. LTFs have been studied since the 1940's, as a model of the "all-or-none character of nervous activity" [MP43]. Understanding the capabilities of collections of LTFs is closely related to understanding the power of neural networks; this connection motivated extensive research on the subject [Mur71]. In the 1960's, Minsky and Papert [MP69] proved many limitations on the abilities of single LTFs.

In this paper, we focus on models that appear to be only mild extensions: depth-two LTF circuits (a.k.a. $\mathsf{LTF} \circ \mathsf{LTF}$ circuits) and depth-three LTF circuits with all weights in $\mathbb{Z} \cap [-\mathrm{poly}(n), \mathrm{poly}(n)]$ (a.k.a. $\mathsf{MAJ} \circ \mathsf{MAJ} \circ \mathsf{MAJ}$, or TC_3^0). $\mathsf{LTF} \circ \mathsf{LTF}$ circuits have a number of LTF gates connected directly to the input variables, and a single output LTF gate which can take in input variables as well as outputs of previously computed LTFs; depth-three LTF circuits are defined similarly. These circuits roughly correspond to neural nets with one or two hidden layers, respectively. Minsky and Papert [MP69] proved that $\mathsf{LTF} \circ \mathsf{LTF}$ of $2^{O(n)}$ gates can compute any Boolean function, but failed to prove a strong impossibility result for these circuits.

Despite considerable study in the complexity of neural networks (see Section 2 for more background), the power of a single hidden layer is still poorly understood: prior to our work, it was open whether every function in *nondeterministic* $2^{O(n)}$ *time* could be computed by $\mathsf{LTF} \circ \mathsf{LTF}$ circuit families (or TC_3^0 families) with $O(n)$ gates, or by $\mathsf{LTF} \circ \mathsf{LTF}$ circuit families with $n^{3/2} \cdot \mathrm{poly}(\log n)$ wires. (Linear-gate lower bounds were proved by several groups in the early 90's; an $n^{3/2}$ wire lower bound was proved in 1993 by Impagliazzo, Paturi, and Saks [IPS93]. See Section 2.)

By results of Allender and Koucky [AK10], in order to separate NC^1 from TC^0, we only need to exhibit a function in NC^1 that does not have $n^{1.1}$ gates for *every* depth $d \geq 2$. That is, the problem of proving super-linear gate lower bounds for all $O(1)$-depth threshold circuits turns out to be as difficult as proving *super-polynomial* lower bounds for $O(1)$-depth threshold circuits. Before we can do that, we have to first prove non-linear gate lower bounds for depth-three circuits.

1.1 Our Results

We prove the first non-trivial super-linear gate lower bounds and super-quadratic wire lower bounds for depth-two threshold circuits ($\mathsf{LTF} \circ \mathsf{LTF}$), and depth-three majority circuits (TC_3^0). Our hard functions have much lower complexity than $\mathsf{NTIME}[2^{O(n)}]$; they are in fact computable in P (even in uniform TC^0 itself).

We start with lower bounds for a linear-time computable function known as *Andreev's function*, which also appears in the best known formula size lower bounds. Our lower bounds extend to the average case, showing that small depth-two threshold circuits cannot compute Andreev's function on more than a $(1/2 + o(1))$-fraction of inputs. To define the function, let us set up some notation. For $k \in \mathbb{N}$, the *k-bit multiplexer function* is defined as $M_{2^k}(x_1, \ldots, x_{2^k}, a_1, \ldots, a_k) := x_{\mathrm{bin}(a_1 \ldots a_k)}$, where $\mathrm{bin} : \{0,1\}^k \to \{1, \ldots, 2^k\}$ converts k-bit strings into positive integers. Let $n = 2 \cdot 2^k$. The *Andreev function* A_n [And87] is defined as:

$$A_n(x, a_{1,1} \ldots a_{1,(2^k/k)}, \ldots \ldots, a_{k,1} \ldots a_{k,(2^k/k)}) = M_{2^k}\left(x, \left(\sum_{j=1}^{(2^k/k)} a_{1,j} \bmod 2\right), \ldots, \left(\sum_{j=1}^{(2^k/k)} a_{k,j} \bmod 2\right)\right),$$

where $x \in \{0,1\}^{2^k}$, and $a_{i,j} \in \{0,1\}$. In words, A_n computes the parity on k disjoint sets of $2^k/k$ inputs, then feeds the resulting k-bit string to the multiplexer function on the remaining 2^k inputs x. (For simplicity we may think of k itself as a power of two, so we do not have to worry about divisibility issues with $2^k/k$.)

Since 1987, the function A_n has been a primary target for formula size lower bounds [And87, IN88, PZ93, Hås98, IMZ12]. The best known explicit size lower bounds for formulas over both the DeMorgan

basis ($n^{3-o(1)}$) and the full binary basis ($n^{2-o(1)}$) are achieved by A_n. Our first result is a non-linear gate lower bound for computing A_n with depth-two threshold circuits:

Theorem 1.1. *Any function f that agrees with A_n on at least a $(1/2 + \varepsilon)$-fraction of inputs for some $\varepsilon \gg \sqrt{\log(n)/n}$ cannot be computed by* $\mathsf{LTF} \circ \mathsf{LTF}$ *circuits with fewer than* $\Omega(\varepsilon^3 n^{3/2}/\log^3(n))$ *gates or fewer than* $\Omega(\varepsilon^3 n^{5/2}/\log^{7/2}(n))$ *wires.*

In contrast with these lower bounds, there are several nice (and somewhat easy) circuit constructions for computing Andreev's function as well:

Theorem 1.2. *The function A_n has (uniform) depth-3* $\mathsf{MAJ} \circ \mathsf{MAJ} \circ \mathsf{MAJ}$ *(i.e.* TC_3^0*) circuits of $O(n)$ gates, (uniform)* $\mathsf{LTF} \circ \mathsf{LTF}$ *circuits of $O(n^3/\log n)$ gates, parity decision trees of depth at most $\log_2(n)$, and (uniform)* $\mathsf{MOD3} \circ \mathsf{MOD2}$ *circuits of $O(n^2)$ gates.*

Hence the lower bounds of Theorem 1.1 establish several average-case complexity hierarchies in the low-depth circuit regime: for example, A_n is computable by depth-three LTF circuits of $O(n)$ gates, but is not computable on a $1/2 + \varepsilon$ fraction of inputs by depth-two LTF circuits of $\varepsilon^3 \cdot n^{3/2-o(1)}$ gates.

As our lower bounds are average-case, we easily obtain some lower bounds on computing A_n with *distributions* of $\mathsf{LTF} \circ \mathsf{LTF}$ circuits. Say that a Boolean function f is computed by an ε-*Approximate Majority of* $\mathsf{LTF} \circ \mathsf{LTF}$ if there is a collection $\mathscr{C}$ of $\mathsf{LTF} \circ \mathsf{LTF}$ circuits such that, for every input x, at least a $1/2 + \varepsilon$ fraction of circuits in $\mathscr{C}$ output the value $f(x)$.

Corollary 1.1. *Every ε-Approximate Majority of* $\mathsf{LTF} \circ \mathsf{LTF}$ *for A_n needs at least* $\Omega(\varepsilon^3 n^{3/2}/\log^3(n))$ *gates and* $\Omega(\varepsilon^3 n^{5/2}/\log^{7/2} n)$ *wires.*

This is a partial step towards lower bounds for *depth-three* circuits composed of MAJORITY gates with negations, i.e. the class TC_3^0. It follows from our distribution results that (for example) Andreev's function has no TC_3^0 circuit of $O(n^{1.1})$ gates where the output gate has fan-in $o(n^{2/15})$.

Onward to Depth Three. However, as stated in Theorem 1.2, Andreev's function has $O(n)$-gate TC_3^0 circuits. To obtain super-linear gate and super-quadratic wire lower bounds in the depth-three setting, we modify Andreev somewhat, defining a new explicit function B_n. Informally, B_n has the same inputs (x, a) as A_n with $|x| = |a|$, and as before the function divides its string a into groups and takes parities of each group, but B_n also feeds x into the generator matrix of an $1/\mathrm{poly}(n)$-balanced error-correcting code (i.e. a $1/\mathrm{poly}(n)$-biased set) before calling the multiplexer. This is similar to a function constructed by Komargodski and Raz [KR13], who also used error-correcting codes in a modification of Andreev's function to prove average-case formula lower bounds. We let $\mathsf{MAJ} \circ \mathsf{LTF} \circ \mathsf{LTF}$ be the class of circuits which compute a majority value of depth-two threshold circuits.

Theorem 1.3. *There is no* $\mathsf{MAJ} \circ \mathsf{LTF} \circ \mathsf{LTF}$ *circuit of $o(n^{3/2}/\log^3 n)$ gates or $o(n^{5/2}/\log^{7/2} n)$ wires that computes B_n.*

Tight Results for PARITY in Depth-Two. Finally, we illustrate the strength of our techniques by proving asymptotically tight results on approximating the PARITY function with $\mathsf{LTF} \circ \mathsf{LTF}$ circuits:

Theorem 1.4. *The gate complexity of* $\mathsf{MAJ} \circ \mathsf{MAJ}$ *circuits that agree with PARITY on* 99% *of all n-bit inputs is $\Theta(\sqrt{n})$. The wire complexity is $\Theta(n^{3/2})$. The lower bounds hold even for* $\mathsf{LTF} \circ \mathsf{LTF}$ *circuits.*

Theorem 1.4 shows that the $\Omega(n^{3/2})$ wire lower bound and $\Omega(n^{1/2})$ gate lower bound for PARITY proved by Impagliazzo, Paturi, and Saks [IPS93] are both tight in the average case.

1.2 Intuition

The key to our lower bounds is a new random restriction lemma for LTFs. With random restrictions, one generally studies the probability that a distribution of partial Boolean assignments to the inputs of a circuit "forces" many gates of that circuit to output a fixed value on all remaining inputs. The idea of forcing is very natural for circuits made of AND and OR gates: an AND is "forced to 0" when one of its inputs is assigned 0, and an OR is "forced to 1" when one of its inputs is 1. As a result, random restrictions have been rather effective for analyzing AC^0 circuits made out of unbounded fan-in AND and OR gates (for example [FSS84, Yao85, Hås86, Ros08]), as well as formulas over the AND/OR/NOT basis (for example [Sub61, And87, PZ93, Hås98]). Strong average-case lower bounds for AC^0 and formula size are also known; some have only been proved very recently [AW85, KR13, IMZ12, KRT13, Hås14, RST15].

For linear threshold functions, the notion of "forcing to a constant" is more subtle. There are two ways we could conclude that a partially restricted LTF is equivalent to a constant function: either the partial Boolean assignment makes part of the LTF so large that the threshold value is achieved on all remaining inputs, or the assignment makes the LTF so small that the threshold value is never achieved. Hence a threshold function can be "forced to 0" in some cases, and 1 in other cases. Also, note that if an LTF only depends non-trivially on a single variable after a restriction (that is, no other variables can affect the output value), then the LTF gate can be removed, and replaced with a single wire coming from that single variable. We incorporate both kinds of reasoning in our arguments.

In order for our analysis to work, we need to consider random restrictions of a structured yet "adversarial" form. In particular, let $\mathscr{P}$ be an *arbitrary* partition of $\{1, 2, \ldots, n\}$ into equal parts, and let $\mathscr{R}_{\mathscr{P}}$ be the distribution of random restrictions on n Boolean variables which randomly fixes all but one element of each part of $\mathscr{P}$. For our applications, one should think of $\mathscr{P}$ as partitioning $[n]$ into k blocks of size n/k for some $k \ll n$, so that $\mathscr{R}_{\mathscr{P}}$ roughly corresponds to fixing all but k randomly chosen variables.

Lemma 1.1 (Random Restrictions on Linear Threshold Functions). *Let $f : \{0,1\}^n \to \{0,1\}$ be a linear threshold function. Let $\mathscr{P}$ be a partition of $[n]$ into parts of equal size, and let $\mathscr{R}_{\mathscr{P}}$ be the distribution on restrictions $\rho : [n] \to \{0, 1, \star\}$ that randomly fixes all but one element of each part of $\mathscr{P}$. Then*

$$\Pr_{\rho \sim \mathscr{R}_{\mathscr{P}}} [f \text{ is not forced to a constant by } \rho] = O(|\mathscr{P}|/\sqrt{n}).$$

For example, given a partition of the inputs with at most $\sqrt{n}/1000$ parts, it is very likely that f is equivalent to a constant function, when a random element of each part is left unrestricted and all other inputs are assigned to random bits.

The primary tool in our random restriction analysis is a well-known result in additive combinatorics by Littlewood and Offord [LO43, Erd45] which (tightly) upper bounds the probability (on a random Boolean assignment) that a linear function takes on a value in a given interval of length 2. We can use this lemma to closely estimate the probability that an LTF is "forced to a constant" when some of its inputs are randomly set to 0 or 1. The central intuition is that a linear threshold function is not "forced to a constant" by a partial Boolean assignment precisely when the restricted part of the linear function lies in a certain integer interval, certifying that the restricted part is neither "too high" to always exceed the threshold, nor "too low" to always fail to meet the threshold. Adapting Littlewood-Offord to this event requires some care, as we need to compute probability upper bounds for (potentially) large intervals defined by the LTF.

Another key idea in our lower bound proofs is the fact that there exist relatively hard functions for low-depth threshold circuits. This is obtained by a counting argument, showing that since there are few distinct functions computed by small threshold circuits, there must be hard functions which cannot be computed by any of them (or any small group of them).

Let us illustrate these two ideas, by sketching the proof of Theorem 1.1. By plugging in the correct bits of a hard function into a multiplexer, we can force our candidate circuit C to compute the value of such a hard function. (This step was already known to Andreev [And87].) This would produce lower bounds of order $\tilde{\Omega}(n)$. However, the circuit C must still be capable of computing the hard function, even after we have restricted it to $O(\log n)$ inputs. Since C must still be large after a restriction has killed all but a $1/\sqrt{n}$ fraction of its gates, it must have originally had size $\tilde{\Omega}(n^{3/2})$, by our Random Restriction Lemma.

Comparison with Impagliazzo, Paturi, and Saks. In the previous best known lower bounds on LTF circuits of depth at least two, Impagliazzo, Paturi, and Saks [IPS93] also used a random restriction method to prove their lower bounds. There are several critical differences between their approach and ours. Their main lemmas state that for any LTF circuit with n variables and δn wires, there is a variable restriction which leaves $\Omega(n/\delta^2)$ variables unset and makes *every* bottom-layer gate dependent on at most one variable. First, observe that their lemmas are nontrivial only when the number of wires is $O(n^{3/2})$. Second, the proofs of their lemmas require that they pick a particular *random* partition of variables in which the restriction is performed; we will need to allow an adversary to control the partition in order for our analysis to work. Third, instead of insisting that every bottom-layer gate is reduced, we use Littlewood-Offord to calculate the probability that single gate is either forced to a constant (Lemma 1.1) or depends on at most one variable (Lemma 3.1). Fourth, we use our relaxed setting to incorporate strong LTF ∘ LTF lower bounds for random functions in our arguments to gain an extra linear factor in the gate and wire lower bounds (Theorem 1.1), and we add another layer of complexity to the function to insert correlation-style arguments to gain an extra layer of circuit depth (Theorem 1.3).

2 Preliminaries

We denote assignments to n Boolean variables by functions of the form $\tau : [n] \to \{0,1\}$.

In our proofs, a *linear threshold function* (LTF) is defined by a pair (L,t) where $t \in \mathbb{R}$ and $L : \{0,1\}^n \to \mathbb{Z}$ is a function of the form $L(x_1,\ldots,x_n) = \sum_{i=1}^{n} a_i x_i$ where all $a_i \in \mathbb{R}$. The *output* of (L,t) on an assignment $\tau : [n] \to \{0,1\}$ is $[\![L(\tau(1),\ldots,\tau(n)) \geq t]\!]$, where $[\![P]\!]$ is notation for the function which outputs 1 when property P is true and 0 when P is false.

The Littlewood-Offord Lemma. We will apply a classical result of Littlewood and Offord [LO43] which upper bounds the number of inputs to a linear function $L(x) = \sum_{i=1}^{t} a_i x_i$ so that the output lies in a given interval I of length 2. In particular, if at least n of the a_i have $|a_i| \geq 1$, and if x is a random point in $\{-1,1\}^t$, then $\Pr_x[L(x) \in I] \leq O(\log(n)/\sqrt{n})$. The bound was later improved to $O(1/\sqrt{n})$ by Erdős [Erd45]. By scaling the length of the interval I, we obtain the following:

Lemma 2.1. *Let $L(x) = \sum_{i=1}^{n} a_i x_i$ be a linear function, and $k \in \mathbb{N}$. Let $I \subset \mathbb{R}$ be a finite interval, and suppose that $|a_i| \geq |I|$ for at least k of the a_i. Then*

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

where $\in_u$ denotes a uniform random choice.

Proof. Start with the original lemma: assume at least k of the a_i have $|a_i| \geq 1$, let I be an arbitrary interval of length 2, and obtain $\Pr_{x \in \{-1,1\}^t}[L(x) \in I] \leq O(1/\sqrt{k})$. What follows is some simple massaging of this statement.

If we add $\sum_{i=1}^{n} a_i$ to the function $L(x)$, then we can let each x_i be chosen from $\{0,2\}$. Then, shifting the interval I by $\sum_{i=1}^{n} a_i$ (defining an interval I' of length 2), we still have $\Pr_{x \in \{0,2\}^n}[L(x) \in I'] \leq O(1/\sqrt{k})$. Dividing everything by $1/2$, the interval I' becomes length 1, and we obtain $\Pr_{x \in \{0,1\}^n}[L(x) \in I'] \leq O(1/\sqrt{k})$. Finally, if we multiply the a_i's and I' by any desired interval length $\ell \in \mathbb{R}^+$, we can accommodate any finite interval, as long as at least k of the a_i's have absolute value at least the interval length. $\qquad \square$

Constant-Depth Threshold Lower Bounds for Random Functions   Another ingredient in our main result is a threshold circuit lower bound for random functions. This follows from a counting argument via a non-trivial upper bound on the number of distinct functions computable with low-depth LTF circuits. The upper bound on the number of possible LTFs has been proved many times; the earliest reference we have found is Winder [Win60] from the 1st Annual FOCS:

Theorem 2.1 ([Win60]). *The number of linear threshold functions on n variables is at most $2^{O(n^2)}$.*

The $2^{O(n^2)}$ upper bound immediately follows from Chow's theorem (from the 2nd Annual FOCS), which characterizes linear threshold functions by their low-degree Fourier coefficients.

Theorem 2.2 (Chow [Cho61]). *Every LTF f on n variables is uniquely determined by its $n+1$ Fourier coefficients $\hat{f}(\emptyset), \hat{f}(1), \ldots, \hat{f}(n)$.*

To see why the $2^{O(n^2)}$ upper bound follows from Theorem 2.2, observe that each Fourier coefficient of a Boolean function can take on at most $O(2^n)$ values, because it is an expectation of a random variable taking values in $\{-1, 1\}$, over a 2^n sample space. Combined with Chow's Theorem, the number of LTFs on n variables is at most $O(2^n)^{n+1} \leq 2^{O(n^2)}$. For a proof, see (for example) O'Donnell and Servedio [OS11], or Knuth ([Knu11], Theorem T) who states the theorem slightly differently (the language of Chow, in fact).

A considerable generalization of Winder's theorem was given by Roychowdhury, Siu, and Orlitsky:

Theorem 2.3 ([ROS94b]). *Let $\mathscr{F} = \{f_1, \ldots, f_s\}$ be a fixed collection of functions of the form $f_i : \{0,1\}^n \to \{0,1\}$. Then there are at most $(2^n+1)^{s+1}$ distinct functions $g : \{0,1\}^n \to \{0,1\}$ of the form*

$$g(x_1, \ldots, x_n) = \left[\!\!\left[\sum_{i=1}^{s} w_i \cdot f_i(x_1, \ldots, x_n) \geq t \right]\!\!\right],$$

where $w_1, \ldots, w_s, t \in \mathbb{R}$ and $[\![P]\!]$ is notation for the function which outputs 1 when property P is true and 0 when P is false. That is, there are $O(2^{ns+s})$ threshold functions over s input functions. As a consequence, the number of depth-2 linear threshold circuits of $s \geq n$ gates and n inputs is at most $2^{O(n^2 s)}$.

For completeness, we give a short self-contained proof. Our proof builds on Knuth's elegant proof of Chow's theorem ([Knu11], Theorem T).

Proof. Let g be an LTF taking s inputs, and let $f_1, \ldots, f_s$ be LTFs on n variables. Let $S(g)$ be the set of $y \in \{0,1\}^s$ such that $g(y) = 1$ and $(f_1(x), \ldots, f_s(x)) = y$ for some $x \in \{0,1\}^n$. Let $\Sigma(g) \in \mathbb{N}^s$ be the sum of all $y_i \in S(g)$ as vectors over the integers. As each entry of $\Sigma(g)$ is an integer in $[0, 2^n]$, note there are $(2^n+1)^s$ possible values for $\Sigma(g)$.

We claim that, if $|S(g)| = |S(h)|$ and $\Sigma(g) = \Sigma(h)$ for two LTFs g and h over the same functions $f_1, \ldots, f_s$, then $g(f_1, \ldots, f_s) = h(f_1, \ldots, f_s)$ as Boolean functions.

It follows from the claim that:

(a) Every depth-two LTF circuit of $s+1$ gates is uniquely determined by $|S(g)| \leq 2^n$, $\Sigma(g)$, and the s LTF gates on the bottom layer. Hence there are at most $(2^n+1)^{s+1}$ threshold functions over s given input functions.

(b) There are at most $2^{O(n^2 \cdot s)}$ distinct depth-two LTF circuits, since there are at most $(2^n+1)^{s+1}$ possible choices for the output gate, and $2^{O(n^2 s)}$ choices for the s LTFs on the bottom layer (by Theorem 2.1).

So let's prove the claim. Suppose $g(f_1,\ldots,f_s) \neq h(f_1,\ldots,f_s)$ as Boolean functions, but $|S(g)| = |S(h)|$ and $\Sigma(g) = \Sigma(h)$. Let $\{y_1,\ldots,y_k\} \subseteq \{0,1\}^s$ be the set of all points in the image of $(f_1,\ldots,f_s) : \{0,1\}^n \to \{0,1\}^s$, such that $g(y_i) = 1$ and $h(y_i) = 0$. By definition, $\{y_1,\ldots,y_k\} = (S(g) \setminus S(h))$. Because $|S(g)| = |S(h)|$, there must also be exactly k points $z_1,\ldots,z_k$ in the image of $(f_1,\ldots,f_s)$ such that $g(z_i) = 0$ and $H(z_i) = 1$; we therefore have $\{z_1,\ldots,z_k\} = (S(h) \setminus S(g))$.

Since $\Sigma(g) = \Sigma(h)$, the total sum of all vectors in $S(g)$ and $S(h)$ are the same, so we must have $\sum_{i=1}^k y_i = \sum_{i=1}^k z_i$, where the sum is componentwise.

Suppose the linear function of $g(y)$ has the form $\sum_i w_i y_i$, and threshold value t. Let $w = (w_1,\ldots,w_s)$. By our definition of y_i and z_i, we have $\langle w, y_i \rangle \geq t$ and $\langle w, z_i \rangle < t$ for all i. Therefore

$$\left\langle w, \left(\sum_{i=1}^k y_i\right) \right\rangle = \sum_i \langle w, y_i \rangle \geq kt > \sum_i \langle w, z_i \rangle = \left\langle w, \left(\sum_{i=1}^k z_i\right) \right\rangle.$$

This is a contradiction, since $\sum_{i=1}^k y_i = \sum_{i=1}^k z_i$ and $k > 0$. □

Combining Theorem 2.3 with a simple counting argument, we obtain:

Corollary 2.1. *For all sufficiently large n, a randomly chosen Boolean function on n variables requires depth-2 linear threshold circuits of size at least $\Omega(2^n/n^2)$, with probability $1 - o(1)$.*

Proof. If we choose a function $f : \{0,1\}^n \to \{0,1\}$ uniformly at random, the probability it has depth-2 threshold circuits of size s is at most $2^{O(n^2 s)}/2^{2^n}$, by Theorem 2.3. For $s \leq o(2^n/n^2)$, this probability is $2^{o(2^n)}/2^n = o(1)$. □

By standard arguments we also have an "inapproximability" refinement of the above corollary:

Corollary 2.2. *For all $\varepsilon \gg \sqrt{n/2^n}$, and all but an ε-fraction of n-bit Boolean functions f, there is no depth-2 linear threshold circuit of size $s \leq o(\varepsilon^2 \cdot 2^n/n^2)$ that agrees with f on more than a $(1/2+\varepsilon)$-fraction of inputs.*

Proof. By Theorem 2.3, the total number of functions computed by such size-s depth-2 circuits is at most $2^{o(\varepsilon^2 2^n)}$. For any such circuit, it agrees with a randomly chosen f on a $(1/2+\varepsilon)$-fraction of inputs with probability $2^{-\Omega(\varepsilon^2 2^n)}$, by standard Chernoff bounds. Taking a union bound over our choice of circuits, we conclude that f does not agree with any size-s depth-2 threshold circuit on a $(1/2+\varepsilon)$-fraction of inputs, with probability at least $1 - \varepsilon$. □

A Short History of Low-Depth Threshold Lower Bounds.   Hajnal *et al.* [HMP$^+$93] proved the first size lower bounds for LTF∘LTF circuits, showing that the *inner product modulo 2* (a.k.a. IP2) requires $2^{\Omega(n)}$ size when the weights of each LTF are small (polynomial in the input length). This result is often cited as saying that the inner product does not have subexponential-size MAJ∘MAJ circuits, as MAJORITY functions can simulate LTFs with polynomial weights. Note that IP2 has MAJ∘MAJ∘MAJ circuits with $O(n)$ gates, so we cannot use IP2 in our depth-three lower bounds. Nisan [Nis94] elegantly applied communication complexity ideas to extend the exponential lower bound to MAJ∘LTF circuits; that is, the lower bound holds even if the weights of the LTFs on the hidden layer are arbitrary. Later, Forster *et al.* [FKL$^+$01] extended the lower bound to LTF∘MAJ circuits, where only the middle (hidden) layer is restricted to have small weights.

In terms of lower bounds for general LTF∘LTF circuits, only a few results are known. Goldmann, Håstad, and Razborov [GHR92] showed that every LTF∘LTF circuit can be efficiently simulated by a MAJ∘MAJ∘MAJ circuit, and computing IP2 with LTF∘LTF requires $\Omega(n/\log n)$ gates. Groeger and G. Turán [GT91, GT93] and Roychowdhury, Orlitsky, and Siu [ROS94a] proved that IP2 has gate complexity $\Theta(n)$ for LTF circuits; their result has no depth restriction. Paturi and Saks [PS90] showed that PARITY requires $\Omega(n/\log^2 n)$ gates for MAJ∘MAJ circuits. Impagliazzo, Paturi, Saks [IPS93] showed that LTF∘ LTF circuits computing PARITY cannot have $o(n^{3/2})$ wires, nor can they have $o(n^{1/2})$ gates. (Their proof in fact gives a lower bound for all constant depths, although the bounds get smaller as the depth increases.) In the uniform setting, Allender and Koucky [AK10] have shown that for every d, there is an $\varepsilon \in (0,1)$ such that the SAT problem cannot be solved by LOGTIME-uniform depth-d LTF circuits with $O(n^{1+\varepsilon})$ wires. Other more recent work on LTF∘LTF includes [AM05, HP10, HP13, IPS13, Wil14, CS15].

In all the above cases, no super-linear gate lower bounds were known, and no quadratic wire lower bounds were known, even for LTF∘LTF. PARITY is well-known to have MAJ∘MAJ circuits of $O(n)$ gates, and we show in Theorem 1.4 that there *are* always MAJ∘MAJ circuits of $O(n^{1/2})$ gates and $O(n^{3/2})$ wires which agree with PARITY on 99% of the inputs, so it is impossible to extend the lower bounds of [PS90, IPS93] for PARITY in the way we seek (for good reason).

3   Random Restrictions on Linear Threshold Functions

We are ready to give our main lemma on random restrictions to linear threshold functions. To properly state it, we need to set up some notation. Define a *restriction* to be a function $\rho : [n] \to \{0,1,\star\}$. (Such a function is also called a partial assignment.) If $\mathscr{P}$ is a set partition of $[n]$, we say that ρ is a *random restriction across $\mathscr{P}$* if ρ is obtained by first uniformly randomly choosing a one element e_i of each part of $\mathscr{P}$, then setting $\rho(e_i) = \star$ for each i, and setting $\rho(j)$ randomly and independently to either 0 or 1 for all other $j \in [n]$.

A *completion of ρ* is simply a function $\tau : [n] \to \{0,1\}$ such that for all i such that $\rho(i) \neq \star$ we have $\tau(i) = \rho(i)$. That is, τ extends the partial assignment ρ to some full assignment on all variables. We say that an LTF $f : \{0,1\}^n \to \{0,1\}$ is *forced to a constant by restriction ρ* if there is a $c \in \{0,1\}$ such that for all completions τ of ρ, the output of f on τ always equals c. That is, f is "forced to a constant" if ρ has set enough variables of f that the remaining function is constant. We record the following trivial (but crucial) observation that we can *simplify* circuits when their gates are forced to constants:

Proposition 1. *Let C be an n-input circuit over LTF gates, let ℓ be an LTF in C, and let $\rho : [n] \to \{0,1,\star\}$ be a restriction that forced ℓ to a constant c. Then the subfunction defined by $C(\rho(1),\ldots,\rho(n))$ has an equivalent circuit with the gate ℓ removed, and the constant c placed on the output wires of ℓ.*

We recall the main lemma to prove:

Reminder of Lemma 1.1 *Let $f : \{0,1\}^n \to \{0,1\}$ be a linear threshold function. Let $\mathscr{P}$ be a partition of $[n]$ into parts of equal size, and let $\mathscr{R}_{\mathscr{P}}$ be the distribution on restrictions $\rho : [n] \to \{0,1,\star\}$ that randomly*

fixes all but one element of each part of $\mathscr{P}$. Then

$$\Pr_{\rho \sim \mathscr{R}_{\mathscr{P}}} [f \text{ is not forced to a constant by } \rho] = O(|\mathscr{P}|/\sqrt{n}).$$

Observe that Lemma 1.1 is essentially tight for the MAJORITY function (the LTF with $L(x) = \sum_{i=1}^{n} x_i$ and $t = \lceil n/2 \rceil$). In particular, if we set all but k inputs of the n-bit MAJORITY function to uniform random values, the MAJORITY function is not forced to a constant only when the difference between the number of 1's and the number of 0's is within $[-k, k]$. This occurs with probability $\sim k/\sqrt{n}$ for small k. (This is just another way of saying that the Littlewood-Offord Lemma is tight for the vector $a = (1, \ldots, 1)$.)

Proof of Lemma 1.1. Let the linear threshold function f be defined by the linear function $L(x) = \sum_{i=1}^{n} a_i x_i$ with threshold value $t \in \mathbb{R}$. For a set $B \subseteq [n]$, let $\mathscr{R}_B$ be the distribution of restrictions of f to all but B, where we set elements in B to $\star$ and elements not in B to random $0, 1$ values. Assume $x \in \{0,1\}^{n-|B|}$ is chosen uniformly at random, and let $L'(x) = \sum_{i \notin B} a_i x_i$. Observe that f is forced to a constant if and only if

$$\left(L'(x) < t - \sum_{i \in B: a_i > 0} |a_i| \right) \text{ or } \left(L'(x) > t + \sum_{i \in B: a_i < 0} |a_i| \right).$$

(In the first case, f is forced to 0; in the second case, f is forced to 1.) Therefore, the probability that a random restriction $\rho \sim \mathscr{R}_B$ does not force f to a constant is at most the probability that $L'(x)$ lies in the interval

$$I = \left[t - \sum_{i \in B: a_i > 0} |a_i|, \ t + \sum_{i \in B: a_i < 0} |a_i| \right].$$

Note that $|I| = \sum_{i \in B} |a_i|$, and that we can write I an a union of intervals I_i with $|I_i| = |a_i|$. Therefore, by a union bound we have that

$$\Pr_{\rho \sim \mathscr{R}_B} [f \text{ is not forced to a constant by } \rho] \leq \sum_{i \in B} \Pr_{x} \left[L'(x) \in I_i \right]. \tag{1}$$

For all $i = 1, \ldots, n$, define k_i to be the number of $j \in [n]$ such that $|a_i| \geq |a_j|$. Observe that

$$k_i \geq k_j \iff |a_i| \leq |a_j|. \tag{2}$$

We claim that

$$\Pr_{\rho} [f \text{ is not forced to a constant by } \rho] \leq \sum_{i \in B} \frac{O(1)}{\sqrt{k_i}}. \tag{3}$$

Let us prove this claim. For all $i = 1, \ldots, n$, define k_i' to be the number of $j \in [n] - B$ such that $|a_i| \geq |a_j|$. By Lemma 2.1, we have for all $i = 1, \ldots, |B|$ that

$$\Pr_{x \in \{0,1\}^{n-|B|}} \left[L'(x) \in I_i \right] \leq \frac{O(1)}{\sqrt{k_i'}}.$$

Obviously, $k_i' \leq k_i$ for all $i \in B$. To prove (3), we need an inequality in the opposite direction. We consider two cases. First, suppose there is an $i \in B$ with at least $k_i/2$ different $j \in B$ satisfying $k_j \leq k_i$. Then $\sum_{i \in B} 1/\sqrt{k_i} \geq (k_i/2)/\sqrt{k_i} \gg 1$, and inequality (3) trivially holds in this case. Otherwise, for all $i \in B$, there are at most $k_i/2$ different $j \in B$ with $k_j \leq k_i$. By (2), this means that $|a_i| \leq |a_j|$ for at most $k_i/2$ different j's, implying that $k_i \leq k_i' + k_i/2$ for all $i \in B$. So we have $k_i \leq 2k_i'$ for all $i \in B$, and

$$\Pr_{\rho \sim \mathscr{R}_B} [f \text{ is not forced to a constant}] \leq \sum_{i \in B} \Pr_{x} \left[L'(x) \in I_i \right] \leq \sum_{i \in B} \frac{O(1)}{\sqrt{k_i'}} \leq \sum_{i \in B} \frac{O(1)}{\sqrt{k_i}},$$

8

and the inequality (3) holds in this case as well. Therefore

$$
\begin{aligned}
\Pr_{\rho \sim \mathscr{R}_{\mathscr{P}}}[f \text{ not forced to constant by } \rho] &= \underset{B}{} \left[\Pr_{\rho \sim \mathscr{R}_B}[f \text{ not forced to constant by } \rho] \right] \\
&\leq \underset{B}{} \left[\sum_{i \in B} \frac{O(1)}{\sqrt{k_i}} \right] & \text{(by (3))} \\
&= \sum_{i=1}^{n} \frac{O(1)}{\sqrt{k_i}} \cdot \Pr[i \in B] \\
&= \frac{|\mathscr{P}|}{n} \cdot \left(\sum_{i=1}^{n} \frac{O(1)}{\sqrt{k_i}} \right) \\
&\leq \frac{|\mathscr{P}|}{n} \cdot \left(\sum_{\ell=1}^{n} \frac{O(1)}{\sqrt{\ell}} \right) & \text{(by def. of } k_i\text{)} \\
&= O(|\mathscr{P}|/\sqrt{n}), & \text{(by } \sum_{k=1}^{n} k^{-1/2} \leq 2n^{1/2}\text{)}
\end{aligned}
$$

where the last inequality follows by upper bounding the sum with an integral. This completes the proof.

$\square$

In order to prove our wire lower bounds, we require a version of Lemma 1.1 that yields better results for thresholds that depend on relatively few inputs. Unfortunately, a threshold gate with a single input has only one wire, and yet has a $|\mathscr{P}|/n$ chance of not being forced to be a constant; this is too weak of a bound for us. On the other hand, we know that a gate with only one input always returns just its input or its negation. In particular, even if a gate is not set to a constant by a restriction, we can still replace it by a single wire if its output depends on only *one* of the remaining inputs. In terms of this event, we can prove a much better probability bound.

For a random restriction ρ and LTF f, we say that $f(x_1, \ldots, x_n)$ is *forced to a function of a single input by* ρ if there is an $i \in [n]$ such that for all completions τ of ρ, the output of f on τ depends only on the variable x_i or is a fixed constant.

Lemma 3.1. *Let $f : \{0,1\}^n \to \{0,1\}$ be a linear threshold function that depends on only w of its inputs. Let $\mathscr{P}$ be a partition of $[n]$ into parts of equal size, and let $\mathscr{R}_{\mathscr{P}}$ be the distribution on restrictions $\rho : [n] \to \{0,1,\star\}$ that randomly fixes all but one element of each part of $\mathscr{P}$. Then*

$$
\Pr_{\rho \sim \mathscr{R}_{\mathscr{P}}}[f \text{ is not forced to a function of a single input by } \rho] = O(w|\mathscr{P}|^{3/2}/n^{3/2}).
$$

Proof. Let S be the set of inputs on which f depends. Define k_i for $i \in S$ as in the proof of Lemma 1.1. Suppose we have a random restriction ρ that fixes all input variables except those in $B \subset [n]$.

We want to upper bound the probability that ρ does not fix f to a function of a single input. In order for this to happen it must be that $|B \cap S| \geq 2$, for otherwise the restriction of f depends only on the coordinates in $S \cap B$. Furthermore, this event happens with probability at most $\sum_{i \in B \cap S} O(1/\sqrt{k_i})$, by Equation (3).

Therefore, the probability that f is not fixed to a function of a single input is

$$
\mathbb{E}_B\left[\llbracket |B \cap S| \geq 2 \rrbracket \cdot \sum_{i \in B \cap S} \frac{O(1)}{\sqrt{k_i}}\right] = \sum_{i \in S} \frac{O(1)}{\sqrt{k_i}} \cdot \Pr[i \in B, |B \cap S| \geq 2]
$$

$$
= \sum_{i \in S} \frac{O(1)}{\sqrt{k_i}} \cdot \min\left\{\frac{|\mathscr{P}|}{n}, \sum_{j \neq i, j \in S} \Pr[i, j \in B]\right\}
$$

$$
= \sum_{i \in S} \frac{O(1)}{\sqrt{k_i}} \cdot \min\left\{\frac{|\mathscr{P}|}{n}, w \cdot \frac{|\mathscr{P}|^2}{n^2}\right\}
$$

$$
\leq \left(w^{1/2} \cdot \frac{|\mathscr{P}|^{3/2}}{n^{3/2}}\right) \sum_{i \in S} \frac{O(1)}{\sqrt{k_i}} \qquad (\min\{a,b\} \leq \sqrt{ab}, \ a,b > 0)
$$

$$
= O(w \cdot |\mathscr{P}|^{3/2}/n^{3/2}). \qquad \left(\text{by } \sum_{i \in S} \frac{1}{\sqrt{k_i}} \leq 2\sqrt{w}\right)
$$

$\square$

4 Tight Upper and Lower Bounds for Approximately Computing PARITY

We begin with the upper and lower bounds for computing PARITY with $\mathsf{LTF} \circ \mathsf{LTF}$ circuits, since the lower bounds here are the easiest example of what our Random Restriction Lemmas can do.

Reminder of Theorem 1.4 *The gate complexity of* $\mathsf{MAJ} \circ \mathsf{MAJ}$ *circuits that agree with PARITY on* 99% *of all n-bit inputs is* $\Theta(\sqrt{n})$. *The wire complexity is* $\Theta(n^{3/2})$. *The lower bounds hold even for* $\mathsf{LTF} \circ \mathsf{LTF}$ *circuits.*

Proof. (Upper bound) We produce an explicit construction. Let $L_k = \llbracket \sum_{i=1}^n x_i \geq k \rrbracket$ be the output of a "threshold-at-least-k" gate over its n inputs. Observe that $L_k - L_{k+1}$ is the 0/1 indicator function of the exact threshold function $\sum_{i=1}^n x_i = k$.

Let $c > 0$ be sufficiently large in the following. We construct the top gate of our depth-two circuit to compute the threshold function defined by

$$
\sum_{k \text{ even}, \ k=n/2-c\sqrt{n}}^{n/2+c\sqrt{n}} L_k - L_{k+1} \geq 1.
$$

This function agrees with parity as long as $n/2 - c\sqrt{n} \leq \sum_{i=1}^n x_i \leq n/2 + c\sqrt{n}$, and for a sufficiently large $c > 0$, this happens on 99% of all inputs. This circuit obviously uses $O(\sqrt{n})$ gates and $O(n^{3/2})$ wires.

(Lower Bound) We begin with the observation (originally due to Minsky and Papert [MP69]) that a single LTF with two inputs cannot approximate PARITY on more than 75% of its inputs. Therefore, for every $\mathsf{LTF} \circ \mathsf{LTF}$ circuit C that approximates PARITY on 99% of all n-bit inputs, if C is randomly restricted on all but two unassigned inputs, then the restricted circuit C' cannot be equivalent to a single LTF with more than 10% probability. On the other hand, the circuit C' will be equivalent to a single LTF gate, unless at least one of the bottom level gates of C' is *not forced to a function on one input*: that is, at least one gate on the bottom layer is neither a projection of one input, nor is it a constant function. Call a gate "trivial" if is forced to a function on one input. Observe that a trivial gate can always be replaced by a single wire, or a constant.

So let C be a depth-2 LTF circuit with s gates and w wires that agrees with PARITY on at least 99% of the n-bit inputs. Let $\mathscr{P}$ be a partition of $[n]$ into two equally sized subsets. Suppose we choose a random restriction ρ from $\mathscr{R}_{\mathscr{P}}$ on the inputs of C, resulting in a restricted circuit C' on two inputs. Our Random Restriction Lemmas 1.1 and 3.1 show that the number of non-trivial bottom level gates of C is, in expectation,

$$O\left(\min\left(\frac{s}{\sqrt{n}}, \frac{w}{n^{3/2}}\right)\right).$$

If either $s \ll \sqrt{n}$ or $w \ll n^{3/2}$, then by Markov's inequality, there are no non-trivial gates on the bottom layer with probability at least 50%. That is, on at least half of the possible random restrictions, the remaining circuit C' on two inputs is equivalent to a single LTF gate, and hence C' does not compute PARITY correctly on more than 75% of its inputs. By the previous paragraph, it follows that C does not compute PARITY correctly on 99% of the inputs. $\square$

5 Depth-Two Lower Bounds for the Andreev Function

The strategy for our depth-two lower bounds for A_n has a similar structure to the known Boolean formula lower bound proofs for A_n: hit the function A_n (and a circuit C that supposedly computes it) with a random restriction of an appropriately controlled form, so that the 2^k-bit input x of A_n is assigned a uniform random value, and from each block i there is one $a_{i,j}$ that is left unset. The remainder implements a function on k bits, whose truth table is given by x. When x is chosen uniformly at random, we are implementing a random k-bit function and can use our depth-two lower bounds for random functions (Corollaries 2.1 and 2.2) to argue that the number of gates left in C must be somewhat large: at least $\Omega(2^k/k^2) = \Omega(n/\log^2 n)$. However, if the original C began with a small enough number of gates, the Random Restriction Lemmas tell us that we should expect the random restriction to C to force many bottom-layer gates to constants (and kill many wires in C). Setting the parameters appropriately yields a contradiction.

Reminder of Theorem 1.1 *Any function f that agrees with A_n on at least a $(1/2+\varepsilon)$-fraction of inputs for some $\varepsilon \gg \sqrt{\log(n)/n}$ cannot be computed by* LTF $\circ$ LTF *circuits with fewer than $\Omega(\varepsilon^3 n^{3/2}/\log^3(n))$ gates or fewer than $\Omega(\varepsilon^3 n^{5/2}/\log^{7/2}(n))$ wires.*

Proof. Recall that A_n has inputs $x \in \{0,1\}^{n/2}$, and $a_{i,j} \in \{0,1\}$, with $i = 1,\ldots,k$ and $j = 1,\ldots,2^k/k$, where $k := \lfloor \log_2(n) \rfloor$. Let C be a LTF $\circ$ LTF circuit with $s \le c\varepsilon^3 n^{3/2}/\log^3(n)$ gates or fewer than $w \le c\varepsilon^3 n^{5/2}/\log^{7/2}(n)$ wires for a sufficiently small constant $c > 0$. By Corollary 2.2 there exists a $c' > 0$ so that with probability at least $1 - \varepsilon/3$ a random function f on $\lfloor \log_2(n/2) \rfloor$ bits does not agree on a $(1/2+\varepsilon/3)$-fraction of inputs with any LTF $\circ$ LTF circuit with fewer than $c'\varepsilon^2 n/\log^2(n)$ bottom level gates. We claim that if a random choice of x corresponds to the truth table of such a function f (and it will, with probability at least $1 - \varepsilon/3$), then the probability over the remaining bits that C agrees with A_n is at most $1/2 + 2\varepsilon/3$.

Consider fixing the bits of input to the coordinates corresponding to x to such an f. Let $\mathscr{P}$ be the partition of the remaining $n/2$ bits of input into the k subsets $\{a_{i,1}, a_{i,2}, \ldots, a_{i,2^k/k}\}$. Let ρ be a random restriction from $\mathscr{R}_{\mathscr{P}}$. By Lemmas 1.1 and 3.1, every linear threshold function g on $n/2$ bits is forced to a constant by ρ with probability at least $1 - O(|\mathscr{P}|/\sqrt{n}) = 1 - O(\log(n)/\sqrt{n})$ and every gate with u wires is forced to a function of a single input with probability at least $1 - O(u|\mathscr{P}|^{3/2}/n^{3/2}) = 1 - O(u\log^{3/2}(n)/n^{3/2})$. Therefore, in either case, the expected number of bottom level gates in C not forced to constants by ρ is at most most $c'\varepsilon^3 n/(3\log^2(n))$ (assuming c was sufficiently small).

Therefore, by Markov's inequality, with probability at least $1 - \varepsilon/3$ we have that when C is restricted by ρ, all but $c'\varepsilon^2 n/\log^2(n)$ of the bottom level gates of C are forced to functions of single inputs. In this case,

C is equivalent to a $\mathsf{LTF} \circ \mathsf{LTF}$ circuit with at most $c'\varepsilon^2 n / \log^2(n)$ gates. On the other hand, the restriction of A_n to ρ is merely the function $f(y_1, y_2, \ldots, y_k)$, where each $(y_i = \sum_j a_{i,j} \bmod 2)$ equals either one of the k unassigned variables, or its negation. (Note that negations of variables cannot change the circuit size: negations can easily be accommodated in the threshold gates.) Since by assumption the function f does not agree with any $\mathsf{LTF} \circ \mathsf{LTF}$ of size $c'\varepsilon^2 n / \log^2(n)$ on more than a $(1/2 + \varepsilon/3)$-fraction of inputs, we find for those values of f and ρ that

$$\Pr_y \left[A_n|_\rho(y) = C|_\rho(y) \right] \leq 1/2 + \varepsilon/3.$$

However, it is easy to see that

$$\Pr_{x \in \{0,1\}^n} [A_n(x) = C(x)] = \mathop{}_{f,\rho}\left[\Pr_y [A_n|_\rho(y) = C|_\rho(y)] \right] \leq 2\varepsilon/3 + (1/2 + \varepsilon/3) \leq 1/2 + \varepsilon.$$

This completes the proof. □

5.1 Approx-MAJ of MAJ of MAJ Lower Bounds for Andreev

From the above lower bounds, it is straightforward to conclude depth-three lower bounds for Andreev's function with an Approximate Majority gate at the top.

Say that a Boolean function f is computed by an *ε-Approximate Majority of* $\mathsf{LTF} \circ \mathsf{LTF}$ if there is a collection of $\mathsf{LTF} \circ \mathsf{LTF}$ circuits such that, for every input x, at least a $1/2 + \varepsilon$ fraction of circuits in the collection output the value $f(x)$. The following is an easy corollary of our depth-two threshold lower bounds for Andreev's function:

Reminder of Corollary 1.1 *Every ε-Approximate Majority of $\mathsf{LTF} \circ \mathsf{LTF}$ for A_n needs at least $\Omega(\varepsilon^3 n^{3/2} / \log^3(n))$ gates and $\Omega(\varepsilon^3 n^{5/2} / \log^{7/2} n)$ wires.*

Proof. Let $\mathscr{C} = \{C_1, \ldots, C_t\}$ be a collection of $\mathsf{LTF} \circ \mathsf{LTF}$ circuits such that on every input x, at least $(1/2 + \varepsilon)t$ of the circuits in $\mathscr{C}$ agree with Andreev's function on x. Assuming the entire collection $\mathscr{C}$ has $o(\varepsilon^3 n^{3/2} / \log^3(n))$ total gates, it follows that every C_i has $o(\varepsilon^3 n^{3/2} / \log^3(n))$ gates (we cannot expect to do much better here, since the C_i's could share almost all of their gates and wires). Let $x_1, \ldots, x_{2^n}$ be a list of all n-bit strings, and form a $t \times 2^n$ Boolean matrix M where $M(i,j) = 1$ if and only if $C_i(x_j)$ equals Andreev's function on x. By Theorem 1.1, each row of M contains less than a $(1/2 + \varepsilon)$-fraction of ones, so the entire matrix has less than a $(1/2 + \varepsilon)$-fraction of ones. However, every column of M contains at least a $(1/2 + \varepsilon)$-fraction of ones, because for every input x, at least $(1/2 + \varepsilon)$ of the circuits in $\mathscr{C}$ correctly compute Andreev's function on x. It follows that the entire matrix has at least a $(1/2 + \varepsilon)$-fraction of ones; this is a contradiction. An analogous argument works for wires, too. □

It follows (for example) that Andreev's function has no TC_3^0 circuit of $O(n^{1.1})$ gates where the output gate has fan-in $o(n^{2/15})$.

5.2 Small Circuits for Andreev's Function

We cannot expect to prove much stronger lower bounds for A_n, as there are nice circuit constructions for the function. In particular, we cannot hope to achieve super-linear gate lower bounds for TC_3^0 using A_n:

Reminder of Theorem 1.2 *The function A_n has (uniform) depth-3 TC^0 circuits of $O(n)$ gates[1], (uniform) $\mathsf{LTF} \circ \mathsf{LTF}$ circuits of $O(n^3 / \log n)$ gates, parity decision trees of depth at most $\log_2(n)$, and (uniform) $\mathsf{MOD3} \circ \mathsf{MOD2}$ circuits of size $O(n^2)$.*

---

[1]Recall that a TC^0 circuit is composed of MAJORITY gates with negations.

Proof. (Sketch) In all below constructions, we use the fact that Andreev's function A_n can be represented straightforwardly as an OR of $n/2$ ANDs of $O(\log n)$ parities over $O(n/\log n)$ variables, where over the entire circuit there are only $O(\log n)$ total parity gates.

1. First, since MAJORITY (a.k.a. MAJ) can simulate AND, we can replace the OR of AND part in the above circuit for A_n with an OR of MAJ; this part has $O(n)$ gates. Each of the $O(\log n)$ different parity functions on $O(n/\log n)$ variables can be computed with a $\mathsf{LTF} \circ \mathsf{MAJ}$ circuit of $O(n/\log n)$ gates, where the weights of the LTF gate are at most polynomial in n, and the sum computed in the top LTF gate always equals either 1 or -1 (indeed, this is true for any symmetric function; see Proposition 1 in [HMP$^+$93]). Given that these LTFs have this property, we can simply "merge" the top part computing an OR-MAJ with the $O(\log n)$ $\mathsf{LTF} \circ \mathsf{MAJ}$ circuits computing the bottom part, resulting in an OR of $\mathsf{MAJ} \circ \mathsf{MAJ}$ circuit of $O(n)$ total gates.

2. Each of the $O(\log n)$-fan-in ANDs in the above circuit for A_n can be computed by a weighted sum of n parities, using the Fourier representation of the $O(\log n)$-bit AND function. The OR of $O(n)$ of these weighted sums can be easily written as an LTF of $O(n^2/\log n)$ parities on $O(n)$ variables. From the previous paragraph, each of these parities can replaced by an $\mathsf{LTF} \circ \mathsf{LTF}$ circuit of $O(n)$ gates. The sums computed in the top LTF gates are either 1 or -1, so these outputs can be composed with the top LTF gate (as in part 1), resulting in an $\mathsf{LTF} \circ \mathsf{LTF}$ circuit of $O(n^3/\log n)$ gates.

3. In $\log_2(n/2)$ depth (and $n/2$ leaves), we can compute each of the $\log_2(n/2)$ parities of A_n one by one, then output the appropriate bit of the multiplexer function with one more layer of depth.

4. Every MOD3-MOD2 circuit C of size s corresponds to some $\mathbb{F}_3$ polynomial $p(x_1,\ldots,x_n)$ of $s-1$ monomials over $\{-1,1\}$, where $C(x) = 1$ if and only if $p(x) = 0 \bmod 3$. Note that, as a polynomial over $\{0,1\}$, the multiplexer function has degree $\log n$ and sparsity n. Hence we can express Andreev's function (namely, a multiplexer of $n/\log n$-length parities) as an $\mathbb{F}_3$-polynomial over $\{-1,1\}$ which is a sum of n products of $\log n$ "terms", where each term has the form $(1+\prod x_i)/2$ and each $\prod x_i$ is over one of the $n/\log n$-length parities in Andreev's function. Expanding this polynomial results in a sum of $O(n^2)$ monomials in total, translating into a MOD3 of $O(n^2)$ MOD2s. $\qquad\square$

6 Depth-Three Lower Bounds

We now turn to proving lower bounds against $\mathsf{MAJ} \circ \mathsf{LTF} \circ \mathsf{LTF}$ circuits computing an explicit function. We begin with the observation that an $s(n)$-size $\mathsf{MAJ} \circ \mathsf{LTF} \circ \mathsf{LTF}$ circuit necessarily has $\Omega(1/s(n))$ correlation with at least one of the $\mathsf{LTF} \circ \mathsf{LTF}$ subcircuits. Therefore, it would suffice to find a function which is not $(1/2 + 1/\mathrm{poly}(n))$-approximable by $\mathsf{LTF} \circ \mathsf{LTF}$ circuits of small size. Unfortunately, Andreev's function is not sufficient for these purposes; it has correlation $\Omega(1/\sqrt{n})$ with the majority over the bits corresponding to x; the problem is essentially that *typical* $\log_2(n)$-bit functions have correlations on the order of $1/\sqrt{n}$ with each other. (This is of course to be expected, since Andreev is in linear-size TC_3^0.)

To overcome this issue, we need to change our hard function. We have to use the $n/2$ bits x of A_n to encode a set of functions on *more than* $\log_2(n)$ bits that *all* have low correlation with each other. We will make use of the following construction:

Proposition 2. *Let $\varepsilon > 0$ and $t,m > 1$. There is a polynomial-time computable matrix $A_{m,t} \in \mathbb{F}_2^{m \times t}$ so that for any $x \neq y$ in $\mathbb{F}_2^t$, we have that Ax and Ay agree on $m(1/2 + O(\varepsilon))$ bits, where $\varepsilon = t/\sqrt{m}$.*

Proof. Alon *et al.* ([AGHP92], Section 4) show how to efficiently construct an (ε')-biased set $S \subseteq \{0,1\}^t$ such that $|S| = ct^2/(\varepsilon')^2$ for some constant $c > 0$. Setting $m = ct^2/(\varepsilon')^2$ means that $\varepsilon' = \sqrt{c}t/\sqrt{m}$; we will

set $\varepsilon = \varepsilon'/\sqrt{c}$. The (ε')-biased property means that for every non-zero vector $v \in \{0,1\}^t$,

$$\Pr_{r \in S}[\langle r, v \rangle \neq 0 \bmod 2] \in ((1/2 - \sqrt{c} \cdot \varepsilon)m, (1/2 + \sqrt{c} \cdot \varepsilon)m).$$

Letting the vectors of S form the rows of $A_{m,t} \in \mathbb{F}_2^{m \times t}$, every column of $A_{m,t}$ has Hamming weight in the interval $((1/2 - \Theta(\varepsilon))m, (1/2 + \Theta(\varepsilon))m)$, the matrix A forms a code of distance $m(1/2 - \Theta(\varepsilon))$, and the conclusion holds. $\square$

Corollary 6.1. *Let m be a power of two, and let $n \leq \sqrt{m}$ be a positive integer. There is a function $F :$ $\{0,1\}^{\log_2(m)} \times \{0,1\}^n \to \{0,1\}$ computable in time $\mathrm{poly}(m,n)$ so that for any strings $x, y \in \{0,1\}^n$ with $x \neq y$ we have $F(z,x) = F(z,y)$ for a $(1/2 + O(\varepsilon))$-fraction of $\log_2(m)$-bit strings z, where $\varepsilon = n/\sqrt{m}$.*

Proof. Set $F(z,x) := M(A_{m,n}x, z)$, where M is the $\log_2(m)$-bit multiplexer, and $A_{m,n} \in \mathbb{F}_2^{m \times n}$ is as in Proposition 2. That is, we apply the linear transformation $A_{m,n}$ to $x \in \{0,1\}^n$ yielding an m-bit string x', then for $z \in \{0,1\}^{\log_2(m)}$ we return the bit in the zth position of x'. $\square$

Just to foreshadow a bit, note that in the construction of our final hard function we will choose m to be a power of two that is roughly n^{16}, so the value of ε will be about $1/n^7$.

Ranging over all possible inputs x, the function F above produces many strings of length m that are nearly uncorrelated with each other. We next prove the somewhat coding-theoretic claim that no particular string can be well-correlated with many of them. Let the *relative Hamming distance* $rh(x,y)$ of x and y be the fraction of bit positions in which x and y agree.

Lemma 6.1. *Let $\mathcal{S}$ be a collection of m-bit strings, any two of which have relative Hamming distance $1/2 + \Omega(\varepsilon)$. Then for any other m-bit string T, there are at most $O(\varepsilon^{-1})$ elements $S \in \mathcal{S}$ such that $|rh(T,S) - 1/2| > \Omega(\varepsilon^{1/2})$.*

Proof. Suppose for sake of contradiction that this is not the case. Then for a sufficiently large constant $C > 0$, there are $S_1, S_2, \ldots, S_t \in \mathcal{S}$ with $t = C\varepsilon^{-1}$, and an m-bit T satisfying $rh(S_i, T) \leq 1/2 - C\sqrt{\varepsilon}$ for all i. For notational convenience, set $S_0 := T$, and construe $S_0, \ldots, S_t$ vectors in $\{\pm 1\}^m$ rather than $\{0,1\}^m$. Consider the $(t+1) \times (t+1)$ Gram matrix G, defined for all $i, j = 0, \ldots, t$ as $G[i,j] = \langle S_i, S_j \rangle / m$. Note that

   (a) G is positive semi-definite,
   (b) $G[i,i] = 1$ for all i,
   (c) $G[0,i] \geq C\sqrt{\varepsilon}$ for all i, and
   (d) $|G[i,j]| \leq O(\varepsilon)$ for all $i \neq j$, $i, j > 0$.

Now define the vector $v = (C^2\varepsilon^{-1/2}, -1, -1, \ldots, -1)$. Observe that

$$\begin{aligned} vGv^T &\leq C^4\varepsilon^{-1} - 2t(C^2\varepsilon^{-1/2})(C\varepsilon^{1/2}) + t + t^2 O(\varepsilon) \\ &= C^4\varepsilon^{-1} - 2C^4\varepsilon^{-1} + C\varepsilon^{-1} + O(C^2)\varepsilon^{-1} \\ &= (-C^4 + O(C^2) + C)\varepsilon^{-1} < 0, \end{aligned}$$

which contradicts the fact that G is positive semi-definite. $\square$

Lemma 6.1 implies that no m-bit string T can have large correlation with $F(-,x)$ for very many strings x. Using the fact that there are relatively few distinct functions computable by small depth-two threshold circuits, we can prove that there must be a string x that agrees with none of them.

14

Corollary 6.2. *For integers n and m with m a power of* 2 *and* $m < 2^{n/2}$, *define F as in Corollary 6.1. There is an* $x \in \{0,1\}^n$ *so that, for all* $\log_2(m)$-*bit functions implementable with* LTF∘LTF *circuits C having* $o(n/\log^2 n)$ *gates,* $C(z)$ *agrees with* $F(z,x)$ *on at most a* $(1/2 + \varepsilon)$ *fraction of* $\log_2(m)$-*bit strings z, where* $\varepsilon = O(n^{1/2})/m^{1/4}$.

Proof. By Lemma 6.1 and Corollary 6.1, every such circuit C agrees with $F(z,x)$ on more than a $(1/2+\varepsilon)$-fraction of z's, for at most $O(m)$ values of x. By Theorem 2.3, there are only $2^{o(n)}$ distinct functions computed by circuits of $o(n/\log^2 n)$ gates. Therefore there are at most $m2^{o(n)}$ strings x so that $F(-,x)$ is $(1/2+\varepsilon)$-approximated by some such circuit. Since this is less than the total number of n-bit strings, there must be some $x \in \{0,1\}^n$ so that $F(-,x)$ is not $(1/2+\varepsilon)$-approximated by any LTF ∘ LTF circuit of $o(n/\log^2 n)$ gates. □

We now introduce our hard function for TC_3^0. It is essentially Andreev's function, but with our F in place of the multiplexer. Earlier, we showed that plugging in a hard function in the x-input of the multiplexer yields a function that is hard to compute by LTF∘LTF circuits (thus giving our lower bound of Theorem 1.1). Now, we want to show that plugging in the "correct" input to F yields a function that is hard to even weakly approximate. Morally speaking, this should produce a function that has less than $1/\text{poly}(n)$ correlation with every small LTF ∘ LTF circuit. Unfortunately, although the intuition of this latter claim breaks down when trying to prove it, the intuition is true enough to yield a MAJ∘LTF∘LTF lower bound.

Definition 6.1 (The Hard Function $B_{n,k}$). *For integers* $k \mid n$, *define the* 2n-*bit function* $B_{n,k}(x,a) := F(z,x)$, *where* $x,a \in \{0,1\}^n$, *and* $z \in \{0,1\}^k$ *is defined by* $z_i = \sum_{j=(n/k)(i-1)+1}^{(n/k)i} a_j \pmod 2$.

That is, we compute PARITY on k groups of n/k variables each from the string a, then feed the resulting k-bit string z into $F(z,x)$. (Recall that $F(z,x)$ itself prints the zth bit of $A_{m,n}x$, where $A_{m,n} \in \mathbb{F}_2^{m \times n}$ is the matrix of an small-biased set on vectors of length n.) Finally, we define $B_n(x,a) := B_{n,k}(x,a)$, where $k = 16\log_2 n$.

Reminder of Theorem 1.3 *There is no* MAJ∘LTF∘LTF *circuit of* $o(n^{3/2}/\log^3 n)$ *gates or* $o(n^{5/2}/\log^{7/2} n)$ *wires that computes* B_n.

Proof. Let $k = 16\log_2 n$ in the following; for convenience we will state lower bounds for $B_{n,k}$ in terms of both n and k. We assume for sake of contradiction that there is some MAJ∘LTF∘LTF circuit C computing $B_{n,k}$, with either $s = o(n^{3/2}/(k\log^2(n)))$ gates or $w = o(n^{5/2}/(k^{3/2}\log^2(n)))$ wires.

We set the input x in $B_{n,k}(x,a)$ to correspond to one of the "bad" strings guaranteed by Corollary 6.2. In particular, no LTF ∘ LTF circuit with k inputs and fewer than $cn/\log^2(n)$ gates agrees with the k-input function $F(-,x)$ on more than a $(1/2 + O(n^{1/2}2^{-k/4}))$-fraction of inputs, for some sufficiently small $c > 0$. Note that $n^{1/2}/2^{k/4} \ll n^{-3}$ by our choice of k.

Next, let $\mathscr{P}$ be the partition of the remaining n bits of input into sets of the form $\{a_{i(n/k)+1}, \ldots, a_{(i+1)(n/k)}\}$ for $0 \le i \le k-1$. By Lemmas 1.1 and 3.1, a random restriction from $\mathscr{R}_{\mathscr{P}}$ has the effect that each bottom level gate has probability $O(k/\sqrt{n})$ of not being fixed by the restriction, and each bottom level gate with u wires leading into it has probability $O(uk^{3/2}/n^{3/2})$ of not being fixed to a function of a single input. Therefore, there is some restriction ρ in this family that leaves at most $O(sk/\sqrt{n}) = o(n/\log^2(n))$ or $O(wk^{3/2}/n^{3/2}) = o(n/\log^2(n))$ bottom level gates which depend non-trivially on more than one input. In either case this is $o(n/\log^2(n))$.

Upon making the restriction ρ, the function $B_{n,k}$ on the remaining k bits is equivalent to the function $z \mapsto F(z,x)$, after possibly negating some of the inputs (note these negations cannot change the circuit size). Our circuit C after the restriction ρ is equivalent to a MAJ∘LTF∘LTF circuit with $o(n/\log^2(n))$ bottom-layer gates and $o(n^{5/2})$ middle-layer gates. As the middle-layer gates correspond to LTF∘LTF circuits with

$o(n/\log^2(n))$ gates, each of their outputs have correlation at most $O(n^{1/2}2^{-k/4}) \ll n^{-3}$ with the output of $B_{n,k}$. However, this means that all of the inputs to the top level majority gate have a $o(1/s)$ correlation with $B_{n,k}$, and thus C and $B_{n,k}$ must disagree on some input. This is a contradiction. $\qquad\square$

Although the above proof suggests that $B_{n,k}$ disagrees with every small $\mathsf{LTF} \circ \mathsf{LTF}$ circuit on more than a $(1/2+\varepsilon)$-fraction of inputs for $\varepsilon = n^{-\omega(1)}$, the reader should note carefully that we do not do this, and we are unable to prove this. (What we prove is that, *after an appropriate restriction*, the *remaining* $\mathsf{LTF} \circ \mathsf{LTF}$ subcircuits have low correlation with the *remaining* subfunction of B_n.) This is because there is always a probability of $\Omega(1/\sqrt{n})$ that no bottom-level gates in our circuit are simplified after restriction, and that C agrees with B_n on all inputs consistent with such restrictions. Therefore, our approach will be insufficient to prove such a correlation bound, for any $\varepsilon < 1/\sqrt{n}$.

7 Conclusion

We conclude with some open questions around linear threshold circuits that appear tractable.

1. Can one obtain asymptotically tight results for computing A_n or our function B_n with low-depth threshold circuits? Can average-case lower bounds be proved with $\varepsilon \leq 1/n^{\omega(1)}$, e.g. if we look at functions in $\mathsf{TIME}[2^{O(n)}]$?

2. Are there polynomial-size $\mathsf{LTF} \circ \mathsf{LTF}$ circuits for the function IP2? Many previous lower bounds on threshold circuits show that IP2 is hard; it is known that $\mathsf{LTF} \circ \mathsf{MAJ}$ and $\mathsf{MAJ} \circ \mathsf{LTF}$ circuits require exponential size to compute IP2. Amano and Maruoka [AM05] point out that any answer to this question, yes or no, would imply new separations of some threshold circuit classes.

3. Is there a faster satisfiability algorithm for $O(n)$-gate $\mathsf{LTF} \circ \mathsf{LTF}$ circuits? Several recent theoretical SAT algorithms are built from lower bound techniques [San10, ST13, CKS14, AWY15, CKK+15]. Currently, non-trivial SAT algorithms are known only for slightly superlinearly many wires [IPS13]. Related to this, it would be interesting if one can prove a "concentration" version of our random restriction lemma, where the number of gates remaining is tightly concentrated around the expectation.

4. The second author [Wil14] has shown that $\mathsf{LTF} \circ \mathsf{LTF}$ circuits with $2^{\mathrm{poly}(n)}$ weights and $2^{\delta n}$ gates (for some fixed $\delta > 0$) can be evaluated on all 2^n Boolean assignments in $2^n \cdot \mathrm{poly}(n)$ time. Can this fast evaluation algorithm be used to prove an exponential gate lower bound?

References

[AGHP92] Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. Simple construction of almost k-wise independent random variables. *Random Struct. Algorithms*, 3(3):289–304, 1992.

[AK10] Eric Allender and Michal Koucký. Amplifying lower bounds by means of self-reducibility. *J. ACM*, 57(3), 2010.

[AM05] Kazuyuki Amano and Akira Maruoka. On the complexity of depth-2 circuits with threshold gates. In *MFCS*, pages 107–118, 2005.

[And87] Alexander E. Andreev. On a method for obtaining more than quadratic effective lower bounds for the complexity of π-schemes. *Moscow Univ. Math. Bull.*, 42:63–66, 1987.

[AW85] Miklós Ajtai and Avi Wigderson. Deterministic simulation of probabilistic constant depth circuits (preliminary version). In *FOCS*, pages 11–19, 1985.

[AWY15]   Amir Abboud, Ryan Williams, and Huacheng Yu. More applications of the polynomial method to algorithm design. In *SODA*, pages 218–230, 2015.

[Cho61]   Chao-Kong Chow. On the characterization of threshold functions. In *SWCT (FOCS)*, pages 34–38, 1961.

[CKK+15]   Ruiwen Chen, Valentine Kabanets, Antonina Kolokolova, Ronen Shaltiel, and David Zuckerman. Mining circuit lower bound proofs for meta-algorithms. *Computational Complexity*, 24(2):333–392, 2015.

[CKS14]   Ruiwen Chen, Valentine Kabanets, and Nitin Saurabh. An improved deterministic #SAT algorithm for small De Morgan formulas. In *MFCS*, pages 165–176, 2014.

[CS15]   Ruiwen Chen and Rahul Santhanam. Improved algorithms for sparse MAX-SAT and MAX-k-CSP. In *Theory and Applications of Satisfiability Testing - SAT*, pages 33–45, 2015.

[Erd45]   P. Erdős. On a lemma of Littlewood and Offord. *Bull. Am. Math. Soc.*, 51:898–902, 1945.

[FKL+01]   Jürgen Forster, Matthias Krause, Satyanarayana V. Lokam, Rustam Mubarakzjanov, Niels Schmitt, and Hans Ulrich Simon. Relations between communication complexity, linear arrangements, and computational complexity. In *FSTTCS 2001: Foundations of Software Technology and Theoretical Computer Science*, pages 171–182. Springer, 2001.

[FSS84]   Merrick L. Furst, James B. Saxe, and Michael Sipser. Parity, circuits, and the polynomial-time hierarchy. *Mathematical Systems Theory*, 17(1):13–27, 1984.

[GHR92]   Mikael Goldmann, Johan Hastad, and Alexander Razborov. Majority gates vs. general weighted threshold gates. *Computational Complexity*, 2:277–300, 1992.

[GT91]   Hans Dietmar Gröger and György Turán. On linear decision trees computing boolean functions. In *ICALP*, pages 707–718, 1991.

[GT93]   Hans Dietmar Gröger and György Turán. A linear lower bound for the size of threshold circuits. *Bull. of the EATCS*, 50:220–221, 1993.

[Hås86]   Johan Håstad. Almost optimal lower bounds for small depth circuits. In *STOC*, pages 6–20, 1986.

[Hås98]   Johan Håstad. The shrinkage exponent of De Morgan formulae is 2. *SIAM J. Comput.*, 27:48–64, 1998.

[Hås14]   Johan Håstad. On the correlation of parity and small-depth circuits. *SIAM J. Comput.*, 43(5):1699–1708, 2014.

[HMP+93]   András Hajnal, Wolfgang Maass, Pavel Pudlák, Mario Szegedy, and György Turán. Threshold circuits of bounded depth. *J. Comput. Syst. Sci.*, 46(2):129–154, 1993.

[HP10]   Kristoffer Arnsfelt Hansen and Vladimir V Podolskii. Exact threshold circuits. In *CCC*, pages 270–279, 2010.

[HP13]   Kristoffer Arnsfelt Hansen and Vladimir V. Podolskii. Polynomial threshold functions and boolean threshold circuits. In *MFCS*, pages 516–527, 2013.

[IMZ12]   Russell Impagliazzo, Raghu Meka, and David Zuckerman. Pseudorandomness from shrinkage. In *FOCS*, pages 111–119, 2012.

[IN88]   Russell Impagliazzo and Moni Naor. Decision trees and downward closures. In *CCC*, pages 29–38, 1988.

[IPS13]   Russell Impagliazzo, Ramamohan Paturi, and Stefan Schneider. A satisfiability algorithm for sparse depth two threshold circuits. In *FOCS*, pages 479–488, 2013.

[IPS93]   Russell Impagliazzo, Ramamohan Paturi, and Michael E. Saks. Size-depth tradeoffs for threshold circuits. *SIAM J. Comput.*, 26(3):693–707, 1997. Preliminary version in STOC'93.

[Knu11]   Donald E. Knuth. *The Art of Computer Programming, Volume 4A: Combinatorial Algorithms, Part 1*. Addison-Wesley Professional, 2011.

[KR13]   Ilan Komargodski and Ran Raz. Average-case lower bounds for formula size. In *STOC*, pages 171–180, 2013.

[KRT13]   Ilan Komargodski, Ran Raz, and Avishay Tal. Improved average-case lower bounds for demorgan formula size. In *FOCS*, pages 588–597, 2013.

[LO43]   J. Littlewood and C. Offord. On the number of real roots of a random algebraic equation III. *Mat. USSR Sb.*, 12:277–285, 1943.

[MP43]   Warren S. McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5(4):115–133, 1943.

[MP69]   Marvin Minsky and Seymour Papert. *Perceptrons: An Introduction to Computational Geometry*. The MIT Press, 1969.

[Mur71]   S. Muroga. *Threshold Logic and its Applications*. John Wiley & Sons, Inc., 1971.

[Nis94]   Noam Nisan. The communication complexity of threshold gates. In *Proceedings of "Combinatorics, Paul Erdos is Eighty"*, pages 301–315, 1994.

[OS11]   Ryan O'Donnell and Rocco A. Servedio. The chow parameters problem. *SIAM J. Comput.*, 40(1):165–199, 2011.

[PS90]   Ramamohan Paturi and Michael E. Saks. Approximating threshold circuits by rational functions. *Inf. Comput.*, 112(2):257–272, 1994. Preliminary version in FOCS'90.

[PZ93]   Mike Paterson and Uri Zwick. Shrinkage of de morgan formulae under restriction. *Random Struct. Algorithms*, 4(2):135–150, 1993.

[ROS94a] Vwani P. Roychowdhury, Alon Orlitsky, and Kai-Yeung Siu. Lower bounds on threshold and related circuits via communication complexity. *IEEE Transactions on Information Theory*, 40(2):467–474, 1994.

[ROS94b] Vwani P. Roychowdhury, Alon Orlitsky, and Kai-Yeung Siu. Neural models and spectral models. In *Theoretical Advances in Neural Computation and Learning*, pages 3–36, 1994.

[Ros08]   Benjamin Rossman. On the constant-depth complexity of k-clique. In *STOC*, pages 721–730, 2008.

[RST15]   Benjamin Rossman, Rocco A. Servedio, and Li-Yang Tan. An average-case depth hierarchy theorem for boolean circuits. In *FOCS*, 2015.

[San10]   Rahul Santhanam. Fighting perebor: New and improved algorithms for formula and QBF satisfiability. In *FOCS*, pages 183–192, 2010.

[ST13]   Kazuhisa Seto and Suguru Tamaki. A satisfiability algorithm and average-case hardness for formulas over the full binary basis. *Computational Complexity*, 22(2):245–274, 2013.

[Sub61]   B. A. Subbotovskaya. Realizations of linear functions by formulas using +,*,-. *Soviet Mathematics Doklady*, 2:110–112, 1961.

[Wil14]   Ryan Williams. New algorithms and lower bounds for circuits with linear threshold gates. In *STOC*, pages 194–202, 2014.

[Win60]   R. O. Winder. *Threshold Logic*. PhD thesis, Princeton University, 1962. Preliminary version in FOCS'60.

[Yao85]   Andrew Chi-Chih Yao. Separating the polynomial-time hierarchy by oracles (preliminary version). In *FOCS*, pages 1–10, 1985.