UNIVERSITY OF CALIFORNIA SAN DIEGO

Using Random Restriction to Prove Lower Bounds for Constant-Depth Threshold Circuits

A thesis submitted in partial satisfaction of the
requirements for the degree Master of Science

in

Computer Science

by

Pawan Charles Paleja

Committee in charge:

Professor Russell Impagliazzo, Chair
Professor Daniel Kane
Professor Ramamohan Paturi

2024

The Thesis of Pawan Charles Paleja is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2024

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would first like to extend my deepest gratitude to Professor Russell Impagliazzo for his guidance and support throughout my time at UCSD. Professor Impagliazzo has been an invaluable part of my education and interest in theoretical computer science from the beginning, when I was merely sitting in on his lab's weekly meetings, all the way to now finishing this thesis. His mentoring has imbued me with the confidence to learn and attempt challenging problems, and I remain extremely grateful for his insights, encouragement and time.

Additionally, this endeavor would not have been possible without Professors Daniel Kane and Ramamohan Paturi. I thank them for their time and expertise and for being part of my thesis committee. Their feedback has greatly enriched the development of my academic endeavors and the development of this thesis. I would also like to express my heartfelt thanks to Professors Joe Politz and Miles Jones for their guidance in my development as an educator, and to all the other remarkable faculty I have had the honor to learn from at UCSD.

Lastly, I would like to give thanks to my friends and family, and especially to my parents, Chandresh and Jagruti. Your belief in me, especially during moments of doubt, has been a constant source of strength. I am profoundly grateful for the love and support which has enabled me to pursue my academic aspirations.

ABSTRACT OF THE THESIS

Using Random Restriction to Prove Lower Bounds for Constant-Depth Threshold Circuits

by

Pawan Charles Paleja

Master of Science in Computer Science

University of California San Diego, 2024

Professor Russell Impagliazzo, Chair

A critical challenge in complexity theory is establishing lower bounds on the size, depth, and complexity of Boolean circuits that compute explicit functions. General Boolean circuits, however, have proven resistant to such lower bounds. Consequently, the community has focused on proving lower bounds for more restricted families of circuits, such as bounded-depth circuits over various bases. A notable success in this area is the use of random restrictions, a method where input variables are fixed according to a probability distribution to simplify the circuit. This thesis explores the application of random restriction techniques to circuits and is structured to provide a thorough understanding of how random restrictions can also be modified and refined to prove more general results.

Chapter 1

Introduction

A longstanding problem in complexity theory is to prove lower bounds on the size, depth, and general complexity of Boolean circuits computing explicit functions in classes of interest (such as arithmetic operations, graph reachability, and satisfiablity, e.g. [10, 6, 2]). However, general boolean circuits seem remarkably resistant to lower bounds of this kind; for example we have been unable to prove that there is a problem in **NP** that requires Boolean circuits over the standard basis AND, OR, NOT computing it to be superlinear in size [15]. Thus, it is quite likely that the community's current techniques are inadequate for this purpose.

It is constructive still, to look at proving explicit lower bounds against more restricted families of circuits (for example bounded-depth circuits over various bases) both in the hopes that the techniques developed may generalize and for their independent interest. Here, the community has had higher levels of success. In particular, random restrictions have emerged as a powerful tool in both classical and recent results. For example, in the aforementioned standard basis, random restrictions were utilized by Ajtai [1], Furst et al [7], Yao [17] and Håstad [8] to show that Parity and Majority function require circuits of exponential size when the depth is bounded.

Plainly put, a random restriction is a process by which some of the input variables of a boolean function (or circuit) are fixed to specific values according to a chosen probability distribution, while the remaining variables are kept free. This process is repeated accordingly until the circuit is simplified to a point that can no longer compute the simplified function. Finally,

some sort of equivalence is reached to show that if the simplified circuit cannot compute the simplified function, then the original circuit cannot compute the original function.

We are interested in the use of this technique when applied to achieve lower bounds against circuits over the basis that includes not only Majority, but can also compute arbitrary threshold functions, that is circuits equipped with unbounded fan-in AND, OR, NOT, THRESHOLD gates. Circuits over this basis are of interest in general both practically as a model for neural networks and theoretically because of their relative strength as a computational model. Indeed, not only can no bounded-depth polynomial-size circuits over the standard basis compute Majority, but such circuits equipped *with* Majority can compute pseudorandom function families which are secure under the hardness of factoring and integer division [13, 9].

This thesis is organized to provide a comprehensive understanding of the use of random restriction techniques in proving lower bounds against Threshold Circuits. We will first begin with a section dedicated to introducing the necessary notation and preliminaries definitions. Following this, we present a simple example of how random restrictions can be applied to demonstrate lower bounds, using the Håstad Switching Lemma as a case study. The subsequent sections are devoted to summarizing and contrasting the approaches and results of applications of the random restriction technique to the class of bounded-depth threshold circuits. We aim specifically to offer a structured and clear exposition of the advancements in random restriction techniques as well as some avenues in how they may be improved.

1.1 Preliminaries

1.1.1 Threshold Circuits, Gates, and Functions

We define a *threshold gate* with fan-in n as a $(n+1)$-tuple $\phi = (\mathbf{w}, \theta)$ where $\mathbf{w} \in \mathbb{R}^n, \theta \in \mathbb{R}$. We call the vector $\mathbf{w}$ the *weights* of ϕ. and θ the *threshold value*. ϕ computes the Boolean function

$$\operatorname{sgn}\left(\langle \mathbf{w}, \mathbf{x} \rangle - \theta\right) = \operatorname{sgn}\left(\sum_{i=1}^{n} w_i x_i - \theta\right).$$

We say that a Boolean function f, represented by some threshold gate f_ϕ, is called a threshold function. We may discuss the inputs to f and ϕ, which we denote as the $\text{supp}(f), \text{supp}(\phi)$. We define the functions $\text{FanIn}(f), \text{FanIn}(\phi) = |\text{supp}(f)|, |\text{supp}(\phi)|$.

Observe that all threshold functions are also *unate*, meaning that they are always either increasing or decreasing in each of the variables. Often, unless there is ambiguity, we will omit ϕ, writing simply f.

In general, a circuit C on n inputs is a directed acyclic graph with an output node and exactly n source nodes, representing the inputs. Each internal node is labeled by a gate with fan-in equal to the in-degree of the node. We will also write C as computing a function $C : \{0,1\} \to \{0,1\}$. Sometimes it will be more favorable for analysis to write C over a different domain, $C : \{-1,1\} \to \{-1,1\}$.

The *gate complexity* of C is equal to the number of non-source nodes of the circuit. The *wire complexity* of C is equal to the number of edges in T. When we refer to a circuit's *size*, we will be referring either to its gate or wire complexity. The *level* of a particular node in C is defined inductively. The source nodes have level 0, while the level of any other node is 1 + the maximum level of its immediate predecessors. The *depth* of C is the level of the output node, and in this thesis, we will be specifically interested where the depth is some constant d. This will also allow us to only consider circuits that are *layered*, that is the inputs to each gate are from gates of one level less, because we can transform any constant depth circuit into a layered circuit by increasing the size by at most d.

We refer to the gates at level ℓ also be the gates at depth $depth - \ell$. For example, we will often discuss gates at the *bottom* level as $\phi_1, ..., \phi_m$, which can equivalently be discussed as the gates at depth d. When disambiguation is unnecessary, we write $f_i = f_{\phi_i}$.

1.1.2 Random Restrictions

For an n variable function or circuit, an *assignment* $\alpha \in \{0,1,*\}^n$ is a n -tuple which represents a mapping of the circuit variables to values, $*$ representing that that variable is kept

unassigned. We will also consider a labelling of $\alpha = (A, U, y)$, where $i \in A$ if $\alpha(x_i) \neq *$, $i \in U$ if $\alpha(x_i) = *$ and $y \in \{0,1\}^{|A|}$ is the bit vector on the assigned values. If $x_1, ...x_n$ are the variables of f then $f(\alpha)$ means that we assign x_i the value of α_i. We will often discuss *partial assignments*, where there is at least one variable kept unassigned, which we will denote in the same way as $f(\alpha)$. This partial assignment means that $f(\alpha)$ is a different threshold function than f, indeed if $A, U \subset [n]$ are indices of the assigned and unassigned variables, then $f(\alpha)$ is the threshold function with weights $w_i, i \in U$ and threshold $\theta' = \theta - \sum_{i \in A} \alpha_i w_i$. We can also discuss applying assignments to gates similarly, which we will notate $\phi|_\alpha$ to denote applying a partial or complete assignment with the same definitions. When discussing *random restrictions*, we will define a space of possible assignments R based on some experiment and sample $\alpha \sim R$. When there is no ambiguity of the experiment we are using, we will simply write α.

We now formally define some random restriction distributions that we will use later.

Definition 1.1 (Blockwise Restriction)**.** Let X be a variable set, and let P be a partitioning of X into $|P|/n$ consecutive blocks. Then define R_P as a distribution on assignments $\alpha : [n] \to \{0,1,*\}$ that randomly fixes all but one element of each part of P.

Definition 1.2 (Uniformly Random Subset)**.** We will use $\mathscr{R}_p^n$ to denote the distribution over restrictions, $\alpha : \{-1,1,*\} \to \{-1,1\}$, such that $\alpha(x) = *$ with probability p and $-1, 1$ with probability $(1-p)/2$ each.

Restriction Trees

It will sometimes be beneficial to consider the space of assignments over a tree. A *restriction tree* T on $\{-1,1\}^n$ of depth h is a binary tree of depth h all of whose internal nodes are labeled by one of n variables, and the outgoing edges from an internal node are labeled $+1$ and -1; we assume that a node and its ancestor never query the same variable. Each leaf ℓ of T defines a restriction α_ℓ that sets all the variables on the path from the root of the decision tree to ℓ and leaves the remaining variables unset. A random restriction tree $\mathscr{T}$ of depth h is a distribution over restriction trees of depth h.

Given a restriction tree T, the process of choosing a random edge out of each internal node generates a distribution over the leaves of the tree (note that this distribution is not uniform: the weight it puts on leaf ℓ at depth d is 2^{-d}). We use the notation $\ell \sim T$ to denote a leaf ℓ of T picked according to this distribution.

A *decision tree* is a restriction tree all of whose leaves are labeled either by $+1$ or -1. We say a decision tree has size s if the tree has s leaves. We say a decision tree computes a function $f : \{-1,1\}^n \to \{-1,1\}$ if for each leaf ℓ of the tree, $f|_{\rho_\ell}$ is equal to the label of ℓ.

1.1.3 Useful Facts

We also state the following bounds that will be useful in our analysis.

Theorem 1.3 (Chernoff Bound)**.** *Let $X_1, X_2, \ldots, X_n$ be independent Bernoulli random variables with $Pr[X_i = 1] = p_i$ for $i = 1, 2, \ldots, n$. Let $X = \sum_{i=1}^{n} X_i$ be their sum. Then, for any $\delta > 0$,*

$$Pr[|X - \mu| \geq \delta\mu] \leq 2\exp\left(-\frac{\delta^2\mu}{3}\right)$$

where $\mu = \mathbb{E}[X] = \sum_{i=1}^{n} p_i$.

Theorem 1.4 (Markov's Inequality)**.** *Let X be a non-negative random variable.*

1. *For any $k > 0$, we have:*
$$\Pr[X \geq k] \leq \frac{\mathbb{E}[X]}{k}.$$

2. *For any $a > 0$, we have:*
$$\Pr[X \geq a\mathbb{E}[X]] \leq \frac{1}{a}.$$

Theorem 1.5 (Chebyshev's Inequality)**.** *Let X be a random variable with finite mean μ and finite non-zero variance σ^2.*

1. *For any $k > 0$, we have:*
$$Pr[|X - \mu| \geq k\sigma] \leq \frac{1}{k^2}.$$

2. *For any $t > 0$, we have:*

$$Pr[|X - \mu| \geq t] \leq \frac{\sigma^2}{t^2}.$$

Fact 1.6. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be a boolean function, then*

$$\mathrm{Var}[f] = 4 \Pr_x[f(x) = 1] \Pr_x[f(x) = -1].$$

Moreover, if

$$p = \min\left(\Pr_x[f(x) = 1], \Pr_x[f(x) = -1]\right).$$

Then, $\mathrm{Var}(f) = \Theta(p)$.

Fact 1.7. *Let $f, g, h : \{-1,1\}^n \to \{-1,1\}$ be arbitrary.*

1. *$Corr(f,g) \in [0,1]$.*

2. *If $Corr(f,g) \leq \varepsilon$ and $\delta(g,h) \leq \delta$, then $Corr(f,h) \leq \varepsilon + 2\delta$.*

3. *Let $\mathscr{T}$ be any random restriction tree. Then*

$$\mathrm{Corr}(f,g) \leq \mathbb{E}_{T \sim \mathscr{T}, \ell \sim T}\left[Corr(f|_{\rho_\ell}, g|_{\rho_\ell})\right].$$

1.2 Warmup: AC0 lower bounds via Håstad Switching Lemma

In the introduction, we alluded to how one of the classical results using a random restriction technique was to prove that bounded depth, unbounded fanin circuits over the standard basis (we will refer to this class now only as AC0) required exponentially many gates to compute the PARITY function. In this section, we provide a warmup to the random restriction techniques we will describe later, by proving this result. First, we state the lower bound formally.

Theorem 1.8. *Any AC0 circuit of depth d computing the parity function on n variables has gate complexity* $s \geq 2^{\Omega(n^{1/d})}$.

We will showcase the random restriction framework to prove this lowerbound. In general, what would want to show is that setting some (but importantly, not all) of the inputs to a given circuit will allow us to greatly simplify it. We then repeat the process of setting some inputs, simplifying the circuit further and further until at some point it becomes to simple to compute the given function. This intution is why Parity can be useful to prove lower bounds, because it is a fact that every restriction of $PARITY_n$ to n_0 variables is still either $PARITY_n$ or its negation. We formalize this into a type of lemma that we will prove throughout this thesis, the depth reduction lemma.

Lemma 1.9 (Depth Reduction of AC0 circuit via Switching Lemma)**.** *Let C be an AC0 circuit computing parity of size S. WLOG, assume the circuit is such that each level alternates AND and OR gates, let the bottom layer of C be AND gates, so that the gates at depth* $d-1$ *consist of DNFs of width at most* $s \coloneqq 20\log S$. *Suppose we apply a random restriction* $\alpha \sim \mathscr{R}_n^p,\ p = 1/(20s)$. *Then* $C|_\alpha$ *can be computed by an equivalent depth* $d-1$ *circuit* C' *over* $n_1 = np = \frac{n}{400\log S}$ *variables.*

Following a simple induction, this immediately implies Theorem 1.8 because we can apply it $d-2$ times to get restrictions $\alpha_1, \alpha_2, \ldots \alpha_{k-2}$. Appending these all together $\alpha_0 = \alpha_1\alpha_2\cdots\alpha_{k-1}$, we have that $C|_{\alpha_0}$ is equivalent to a decision tree on

$$n_0 \coloneqq \frac{n}{(400\log S)^{k-2}} \text{ variables.}$$

As stated previously, every restriction of Parity is either Parity or its negation. Thus, $PARITY_{n_0}$ variables requires S to be at least 2^{n_0-1}, proving Theorem 1.8.

The proof of this statement is achieved immediately following what we will refer to in this thesis as a *random restriction lemma*. We state the one for this section below.

Lemma 1.10 (Håstad's Switching Lemma, [8]). *Let f be a CNF/DNF formula on n variables, where each clause has width at most s. Let $\alpha \sim \mathscr{R}_p^n$ be a random restriction that leaves each variable unset with probability p and sets it to 0 or 1 with probability $(1-p)/2$ each. Then for any integer $t \geq 0$,*

$$\Pr_{\alpha \sim \mathscr{R}_p^n}[f(\alpha) \text{ can be computed by a decision tree of depth } t] \leq (5ps)^t.$$

In other words, if we set some of the input bits to a k-CNF randomly, then actually with *exponentially* high probability, the restricted formula $f|_\alpha)$ is much simpler, i.e. a function that is decided by just a few variables. We can use this lemma to prove Theorem 1.8 by using the fact that the gates at depth $d-1$ are by definition a CNF/DNF formula on the inputs. Thus, if we assume that an AC0 circuit exists that computes parity, then interatively set some of its input bits at random, the switching lemma allows us to collapse the bottom two layers into a single layer. Formalizing this intuition gives one a proof of Lemma 1.9 and therefore Theorem 1.8.

Intuition behind the proof of Lemma 1.10

Unfortunately, the formal proof of Lemma 1.10 requires notation and methodology that does not fit into the scope of what we are concerned with. However, the *intuition* behind the proof does give a launchpad for which we can discuss the results found later in this thesis. Indeed, we might ask, how can a random restriction simplify a DNF formula as stated in Håstad's Switching Lemma? Suppose we have a function f computable by a DNF formula of width w. Let $\alpha \sim \mathscr{R}_n^p$ with $p = o(1/w)$. Let T be a term of the DNF. What we can show is that this term is very likely to become simplified by the values set under the random restriction. Indeed, we can look at the cases of what will happen to the variables of T.

1. One of its literals is set to 0. This immediately allows us to delete T from the DNF, so we successfully simplified the term and DNF. We observe that this is highly likely as well, as our restriction chose uniformly from $\{0,1\}$.

2. All of its literals are set to 1. This in fact reduces the entire DNF to output 1.

3. At least of T's literals is kept free, but all the fixed literals are set to 1. However, given that we chose $p = o(1/w)$, there will not be many gates that have this occur, as nearly all their literals will be set.

Intuitively, this would mean that f is highly likely to become constant, i.e. a decision tree on a single variable, which is implied by the Switching Lemma. To prove this and the general lemma, however requires slightly more effort, so we refer the interested reader to [3] which gives a simpler proof than Håstad.

1.3 Can we achieve the same bounds on threshold circuits?

In this section, we were able to understand how we might use the random restriction technique to prove circuit lower bounds. Before we continue on to discuss threshold circuits, it is worth considering if we should even hope to achieve similar bounds for threshold circuits as we could for AC0.

Recall that when we used the random restriction technique to prove that PARITY $\notin$ AC0, we relied on the fact that an AND or OR gate can be reduced to a constant by setting any one variable to the correct value. However, a threshold gate is not "brittle" in the same sense. A threshold function computing the MAJORITY function, for example, needs at least half of its inputs to be set for it to be forced to a constant. It turns out this translates to a drastic increase in strength for this model. It turns out there are explicit constructions for the Parity function using a relatively small number of gates. We state the exact result below.

Theorem 1.11 (Upper Bound on Gate Complexity to compute Parity, Theorem 1 of [16])**.** *Let $d \leq \log(n)$. Then there exists a depth $d+1$ threshold circuit with $O(dn^{1/d})$ gates and $O((dn^{1/d})^2)$ wires that computes $PARITY_n$.*

Despite this, we would still like to know if we can prove lower bounds that match

this upper bound. What we will show in the next section is an attempt to do so which comes marginally close.

Chapter 2

Overcoming The Robustness of Threshold Gates

In this chapter we address the challenges mentioned previously, namely how can we overcome the relative robustness of threshold gates against random restriction techniques. Using, some relatively straightforward observations, we can prove a random restriction lemma that will lead to the following theorem:

Theorem 2.1. *Any threshold circuit of depth d that computes $PARITY_n$ requires at least $O(n/2)^{1/2(d-1)}$ gates.*

Observe that this bound also applies to the number of wires in the circuit as a circuit of the kind we defined that has k gates must have at least $k+1$ wires. However, it turns out we can improve upon this trivial derivation, which we will show later in the section.

2.1 A partitioning random restriction

The random restriction theorem we would like to prove is as follows:

Lemma 2.2 (Theorem 2 from [11])**.** *Let $F = \{f_1, ..., f_m\}$ be collection of m threshold functions on n variables. Then there exists a partial assignment α that leaves at least $\lfloor n/\left(|F|^2+1\right)\rfloor$ variables free such that $\forall f \in F$, $f(\alpha)$ is a constant function.*

As previously stated the collection of threshold functions represents the gates at the bottom level of a threshold circuit. The consequence then is that we can recursively apply Theorem 2.2 to simply an entire circuit to a constant. Formally:

Corollary 2.3 (Corollary 3 from [11]). *Let C be a depth d circuit on n inputs consisting of at most N threshold function gates. Then there exists a partial assignment α leaving $\lfloor n/2N^{2(d-1)} \rfloor$ variables free such that $f_C(\alpha)$ is constant.*

The proof is a simple induction utilizing Lemma 2.2 which we leave to the reader. Corollary 2.3 immediately gives Theorem 2.1 because the only partial assignments that compute the PARITY function are the assignments that assign values to every variable (i.e. the total assignments). Thus, by Corollary 2.3, the number of gates, N satisfies

$$\begin{aligned} 2N^{2(d-1)}/n &\geq 1 \\ 2N^{2(d-1)} &\geq n \\ N &\geq (n/2)^{2(d-1)}. \end{aligned}$$

The rest of the section will be dedicated to proving Lemma 2.4. We start with an intuition. As we said earlier, a majority gate can be forced to a constant by an assignment on $\lceil n/2 \rceil$ of it's variables. In fact this upper bound extends to all threshold functions.

Lemma 2.4 (Lemma 3.2 from [11]). *Let $\phi = (\boldsymbol{w}, \theta)$ be a nonconstant threshold gate on variable set X. Then for the assignment $\beta : \beta_i = \operatorname{sgn}(w_i)$ there exists a $\ell \in \{0, 1, ..., n\}$ such that $f_\phi(\beta_{\leq \ell}) = 0$ and $f_\phi(\overline{\beta}_{\geq \ell}) = 1$.*

This is easy to verify, so we do not prove it here. The immediate result is that any threshold function on n variables has a partial assignment α that leaves at least $\lfloor n/2 \rfloor$ variables free but forces the function to be a constant. The idea is that we want to give a procedure which will pick the "correct" assignment for all the gates at the bottom layer simultaneously.

A_{11}	A_{12}	...	A_{1m}	$\leftarrow$ With f_1
A_{21}	A_{22}	...	A_{2m}	$\leftarrow$ With f_2
.	.	.	.	
A_{m1}	A_{m2}	...	A_{mm}	$\leftarrow$ With f_m

Figure 2.1. Assigned Variable Partitioning

A_{h1}	A_{h2}	...	A_{hj}	$\overline{A_{hj+1}}$	...	$\overline{A_{hm}}$
$\text{sgn}(x)$	$\text{sgn}(x)$	...	$\text{sgn}(x)$	$\overline{\text{sgn}(x)}$	...	$\overline{\text{sgn}(x)}$

Figure 2.2. Division of a particular row f_h

What our procedure will do is arbitrarily partition the variable set into m^2+1 equal blocks. We will randomly pick m^2 of these blocks, A to be the variables that end up with assigned values, while one block U, will be the unassigned variables. Refer to Figure 2.1 for a visualization of the blocks that comprise A, *after the random shuffling*. We can think of blocks in the ith row of A as being randomly allocated to the ith function of F and we will randomly assign the variables of these blocks in such a way to try to make f_i constant.

Recalling Lemma 2.4, for a specific function $f_h = \phi_h = (\mathbf{w_h}, \theta_\mathbf{h})$ (and row A_h) what we will end up doing is "guessing" at the correct value for j using random restriction, setting one side of the row to the values that we think will make it 1 and the other side to the values that we think will make it 0 (see Figure 2.2 for an example). Note that we do not have any way of knowing if the variables in the ith are actually inputs to f_i, but this does not matter. What we want is that our guesses should have some positive probability at simplifying the gate.

We now formally give the procedure that will give the random assignment α we will use to prove Lemma 2.2.

1. Partition X, the variable set of F into $q = m^2+1$ equal blocks.

2. Choose uniformly at random a 1-1 function mapping the q blocks to the entries of $A \cup U$.

3. Choose uniformly at random a vector $(t_1, t_2, ..., t_m$ from the set $\{0,1,2,...,m\}^m$.

4. For each row A_i, generate an assignment α fix the variables according the following rule:

(a) If $x_k \in A_{ij}$ for $j \leq t_i$ assign $a_k = \text{sgn}((w_i)_k)$

(b) If $x_k \in A_{ij}$ for $j > t_i$ assign $a_k = \overline{\text{sgn}((w_i)_k)}$

Lemma 2.5. *For each $h \in [m]$, the probability that $f_h(a)$ is not constant is at most $1/(m+1)$.*

This lemma is sufficient to prove Lemma 2.2 because we can simply union bound over all the gates at the bottom layer. Indeed, since we have m functions in F, the probability that one of them is not constant is at most $m/m+1$. A probabilistic argument implies the existence of an assignment that makes the theorem hold true, as required. It therefore remains to prove Lemma 2.5.

Proof of Lemma 2.5. For the analysis, it will be convenient to assume that after the blocks A and U are assigned and t_h is chosen, we actually make A_{ht_h} the unallocated block and let U be essentially A_{hm+1}. This does not change the distribution of our random procedure, because each gate is still allocated m random sets of variables, with one random set of variables remaining unallocated, and the way the gates' variables are assigned remains unchanged. For the variables not in the blocks assigned to f_h, assume we set them according to the procedure for the other functions $i \neq h$ with the assignment β. Denote $g = f_h(\beta)$. Recall then for $x_k \in A_{hj}$ for $j < t_h$, $j > t_h$, we set $x_k = \text{sgn}((w_h)_k)$ and $x_k = \overline{\text{sgn}((w_i)_k)}$ respectively. By Lemma 2.4, there exists an index ℓ, where if we have an assignment

$$\alpha_k = \begin{cases} \text{sgn}(w_h)_k, k \leq \ell \\ \overline{\text{sgn}(w_h)_k}, k \geq \ell \end{cases},$$

then $g(\alpha)$ is constant. The bad case would be if x_p was in A_{ht_h}, because as we said, we did not assign variables from this block. However if $\ell > t_h$ or $\ell < t_h$, then by definition, the gate becomes constant as we want. Since t_h was chosen uniformly at random, it has probability $1/(m+1)$ of being the one containing ℓ, giving the required probability. This proves Lemma 2.5 and therefore Lemma 2.2. □

2.2 Refining the basic random restriction method

Previously we alluded to how we could do better than the trivial lower bound on wire complexity that would come from Theorem 2.1. We now state this result

Theorem 2.6 (Corollary 2 from [11]). *Let C be a depth d threshold circuit that computes* PARITY$_n$. *Then C has at least $n^{1+\frac{1}{3^{d-1}-1}}/3\sqrt{2}$ edges.*

The improvement is derived from a refinement to the methodology seen in the previous section. Observe that in the proof of the random restriction lemma, we showed that the probability under our random assignment that a given function out of $|F|$ was not constant was at most $1/(|F|+1)$. However, what about the case where we were one set variable away, two set variables away, etc. from the gate being constant? Now previously, it would be quite difficult to quantify this in general, but when we limit the number of wires, in total, that a circuit can have, the number of wires that can be allocated per gate must can be quantified. In effect, while we will do a similar random partition of the variable set, setting variables similar to how we did in Section 2.1, under this observation we will show that for the gates that our restriction fails to simplify, we can actually set a small number of their inputs to complete their simplification.

We state the random restriction lemma now.

Lemma 2.7 (Lemma 3.1 from [11]). *Let F be a collection of m threshold functions on n variables and let $\delta = \frac{1}{n}\sum_{f\in F} s(f)$. Then there exists a partial assignment α that leaves at least $\frac{n}{4\delta^2+2}$ variables free such that for every $f \in F$, $f(\alpha)$ depends on at most one variable.*

We remark that this lemma is different than Lemma 2.4 because we consider the restricted function to be simplified if it is only on a single variable. However, such a gate must by definition output either the value that the input takes or its negation, so the gates at the next level above depend only on the original inputs.

Recall that when proving the lower bound on gate complexity we partitioned the variable set into associations with a particular function in the collection F. We will do something similar

to prove Lemma 2.7, except this time we will not allow x to be associated with a function that does not take x as input. To that end, denote

$$D_x \coloneqq \{f \in F : x \in \mathrm{Supp}(f)\}, \delta_x \coloneqq |D_x|.$$

The other steps of the procedure are similar as well. After partitioning the variable set, we will assign the variables in each block a value randomly to try and achieve the assignment that will neutralize the gate. The final step is to "clean up" any remaining gates that did not become constant/single input by manually setting all the inputs to such gates. We describe the procedure $\mathrm{PROC}(L)$ with parameter L formally below:

1. Partition the variables. Construct a random partition of the variable set, this time into $m+1$ blocks: $C_1, C_2, ..., C_m$ for each function and R as the reserve set of variables that will remain unassigned (for now). We assign x as follows:

$$\Pr[x \in R] = \frac{1}{1+L\delta_x},$$
$$\text{For } f_i \in D_x,\ \Pr[x \in C_i] = \frac{L}{1+L\delta_x}.$$

   Note that if x is not in R, this equates to choosing C_i uniformly from the set of functions x is an input to. The parameter L can be thought of as a sort of weighting *towards/against* putting a variable x in the reserve set R, and will be optimized for later.

2. For each $i \in [m]$, fix all the variables in C_i according to the following rule (done independently for each $i \in [m]$:

   (a) Choose b_i from $\{0, 1, ..., |C_i|\}$.

   (b) Choose a subset B_i uniformly from all b_i-element subsets of C_i. Generate a partial assignment γ_i on the variables of C_i as follows:

      i. If $x_k \in B_i$, assign $\gamma_i(k) = \mathrm{sgn}((w_i)_k)$

ii. If $x_k \notin B_i$, assign $\gamma_i(k) = \overline{\mathrm{sgn}((w_i)_k)}$

(c) Let γ be the union of the partial assignments $\gamma_i, i \in [m]$.

3. Fix some of the variables in R. For each $i \in [m]$, let T_i denote the set of variables on which $f_i(\gamma)$ depends. If $|T_i| > 1$ fix an arbitrary subset $T_i' \subset T_i$ of size $|T_i'| = T_i - 1$. Let α be obtained from γ by setting all the elements of each T_i' to 1.

The third step clearly ensures that no matter what happens in step 2, the partial assignment α is such that $f_i(\alpha)$ depends on at most one variable for each i. The crux of the proof then is to ensure that the total number of variables V that we need to set is not too high. By our construction, we have that

$$V \geq |R| - \sum_{i=1}^{m} \max(0, |T_i| - 1). \tag{2.1}$$

We can easily calculate that

$$\mathbb{E}[|R|] = \sum_{x \in X} \frac{1}{L\delta_x + 1}.$$

What is more difficult is calculating $\mathbb{E}[\max(0, |T_i| - 1)]$. We will show a bound on this expectation. For convenience we will denote $rm(a) := \max(0, a - 1)$.

Lemma 2.8 (Lemma 5.1 from [11])**.** *For each $i \in [m]$,*

$$\mathbb{E}[rm(|T_i|)] \leq \frac{1}{L} \sum_{x \in S(f_i)} \frac{1}{L\delta_x + 1}.$$

Once we have the expectation, Lemma 2.7 quickly follows,

$$\begin{aligned}
\mathbb{E}[V] &\geq \mathbb{E}[R] - \sum_{i=1}^{m} \mathbb{E}[rm(|\tilde{T}_i|)] \\
&\geq \sum_{x \in X} \frac{1}{L\delta_x + 1} - \sum_{i=1}^{m} \frac{1}{L} \sum_{x \in S(f_i)} \frac{1}{L\delta_x + 1} && \text{(By definition and Lemma 2.8} \\
&\geq \sum_{x \in X} \frac{1}{L\delta_x + 1} - \frac{1}{L} \sum_{i=1}^{m} \sum_{x \in X, f_i \in D_x} \frac{1}{L\delta_x + 1} && \text{(By definition of } S(f_i)) \\
&\geq \sum_{x \in X} \frac{1}{L\delta_x + 1} - \frac{1}{L} \sum_{x \in X} \frac{\delta_x}{L\delta_x + 1} && \text{(By definition of } \delta_x) \\
&= \sum_{x \in X} \frac{1 - \delta_x / L}{L\delta_x + 1} \\
&\geq n \left(1 - \frac{\delta / L}{L\delta + 1} \right) && \text{(By Jensen's Inequality)} \\
&\geq \frac{n}{4\delta^2 + 2}. && \text{(Choosing } L = 2\delta)
\end{aligned}$$

Thus, it remains to show Lemma 2.8.

Proof of Lemma 2.8. Fix $h \in [m]$. We would like to analyze $rm(|T_i|)$. Let $U_h = R \cap \text{Supp}(f_h)$ be the set of inputs of f_h that were kept in reserve (i.e. unassigned). Define a random variable

$$\chi_h = \begin{cases} 1, & \text{if } f_h(\gamma) \text{ not constant} \\ 0, & \text{if } f_h(\gamma) \text{ is constant} \end{cases}.$$

Clearly we have that $rm(|T_i|) \leq \chi_h rm(|U_h|)$. We would like to show

$$\mathbb{E}[\chi_h rm(|U_h)] \leq \frac{1}{L} \mathbb{E}[rm(|U_h| + 1]. \tag{2.2}$$

as it would immediately prove Lemma 2.8. We sketch the analysis of (2.2).

Sketch of (2.2). First, denote $K = C_h \cup U_h$, i.e. all the variables our procedure set that were dependent on association with f_h. Similar to the proof of Lemma 2.5, we observe that we

can condition on the event D of a particular instantiation of $C_i, B_i, i \neq h$ to make our analysis easier. By definition, for $i \neq h$, D determines γ_i on the rest of the variables of $\text{Supp}(f_h) - K$. Let $g = f(\gamma_i), i \neq h$ be the function that results from setting those variables according to D. Our goal then becomes

$$\begin{aligned}\mathbb{E}[\chi_h rm(|U_h|)] &\leq \frac{1}{L}\mathbb{E}[rm(|U_h|+1)|D] \\ &\leq \sum_{i=0}^{k} rm(i)\Pr[|U_h| = i|D]\Pr[g(\gamma) \text{ is not constant}|D \wedge (|U_h| = i)]. \end{aligned} \tag{2.4}$$

Because $C_i, i \neq h$ are determined already by D, we have that $|U_h|$ follows a binomial distribution. Indeed, $x \in R$ now with probability $p = \frac{1}{L+1}$ and is in C_h with probability $\frac{L}{L+1} = 1 - p$. Thus we can replace (2.4) with

$$\sum_{i=0}^{k} rm(i)\binom{k}{i}p^i(1-p)^{k-i}\Pr[g(\gamma) \text{ is not constant}|D \wedge (|U_h| = i)]. \tag{2.5}$$

It remains to upper bound $\Pr[g(\gamma) \text{ is not constant}|D \wedge (|U_h| = i)]$. It is instructive to recall some definitions here. Recall that $g(\gamma)$ is now entirely dependent on C_h, B_h after conditioning on D. So we want to understand how C_h, B_h are constructed. We are given that $|U_h| = i$, so this implies that C_h can be thought of as a uniformly chosen $|K| - i$ subset of K, b_h is uniformly chosen element from $\{0, 1, ..., k-i\}$, and B_h is a randomly chosen subset of C_h.

However, it is equivalent to understand B_h, C_h using the following experiment:

1. Choose a random ordering Γ of K.
2. Choose b_h u.a.r from $\{0, 1, ..., k-i\}$.
3. Let B_h be the first b_h elements of K and $\overline{B_h}$ be the last $k - i - b_h$ elements of K.
4. As normal, let $C_h = B_h \cup \overline{B_h}$.

Note that again Lemma 2.4 yields an index j such that on the order Γ of K we have an ℓ where if

we have an assignment

$$\beta_k = \begin{cases} \operatorname{sgn}(w_h)_k, k \leq \ell \\ \overline{\operatorname{sgn}(w_h)_k}, k \geq \ell \end{cases},$$

then $g(\beta)$ is constant. Thus, can verify that $g(\gamma)$ is nonconstant only if b_h satisfies $j-i \leq b_h \leq j-1$. Since b_h is chosen uniformly, this happens with probability $\frac{i}{k-i+1}$. Plugging into (2.5), we have

$$\begin{aligned}
\mathbb{E}[\chi_h rm(|U_h|) \mid D] &\leq \sum_{i=1}^{k} rm(i) \binom{k}{i} p^i (1-p)^{k-i} \frac{i}{k-i+1} \\
&= \frac{p}{1-p} \sum_{i=1}^{k} rm(i) \binom{k}{i-1} p^{i-1} (1-p)^{k-(i-1)} \\
&= \frac{p}{1-p} \sum_{i'=0}^{k-1} rm(i'+1) \binom{k}{i'} p^{i'} (1-p)^{k-i'} \\
&= \frac{p}{1-p} \sum_{i'=0}^{k-1} rm(i'+1) \mathbb{P}[|U_h| = i' \mid D] \\
&\leq \frac{p}{1-p} \mathbb{E}[rm(|U_h|+1) \mid D] \\
&= \frac{1}{L} \mathbb{E}[rm(|U_h|+1) \mid D],
\end{aligned}$$

which completes the proof of Lemma 2.8 and consequently, Lemma 2.7. □

2.3 Drawbacks and challenges

The results we proved in this section are actually the strongest lower bounds that exist for the size of constant depth threshold circuits computing parity. However, it is not without its limitations. One that we can observe immediately is the means by which we simplified the circuit, i.e. the process we used to generate the assignment. We note that in many instances across both results, we only asked for the existence of an assignment that simplified the circuit. This could imply that threshold function may still be able to compute the parity function on many inputs that are not so adversarially chosen. In the next section, we formalize this question

and exposit two results relating to it.

Chapter 3

Average-Case Lower Bounds Through Anti-Concentration Results

In this section, we would first like to expand on the motivation we ended the previous section on. Recall that the lower bounds we showed previously were worst-case lower bounds, that is we showed for any sequence of constant depth threshold circuits, how to construct an input that they would be unable to compute parity on. We would be interested in asking then, do there exist circuits which can *approximately* compute parity? That is

Question 3.1. Let C be a depth d circuit that computes parity on $1-\varepsilon$ inputs. What is the order of the gate complexity of C? What is the order of the wire complexity of C?

Better explicit construction actually exist for this problem than those that exactly match parity, as was shown in Chapter 2. Indeed,

Theorem 3.2 (Theorem 7 in [16])**.** *Let $\varepsilon > 0$ be an arbitrary constant. Then, there is a threshold circuit of depth $O(d)$ $\dot{(}n\log(1/\varepsilon))^{1/(2d-1)}$ gates that computes $PARITY_n$ correctly on $1-\varepsilon$ fraction of inputs.*

Note that this upper bound for approximation nearly matches the lower bound given in the last chapter by Theorem 2.1. Like in the previous section, we would like to ask if we can show a matching lower bound to this upper bound. It turns out that in the case of depth $d = 2$, we actually can.

Theorem 3.3 (Theorem 1.4 from [12]). *Let T be a threshold circuit of depth 2 computing PARITY on 99% of all n-bit inputs. Then T has at least $\Omega(n)$ gates and $\Omega(n^{3/2})$ wires.*

We will spend the first part of this section explaining again how a random restriction technique can be used to prove this theorem. We will then move on to a discussion on challenges to extending such techniques to depths beyond 2, and end the chapter with another result that addresses some of those challenges.

We will split Theorem 3.3 into two sections, dealing with the gate and wire lower bounds separately. We begin with the gate lower bound.

3.1 Correlation bounds in depth 2

We first restate the relevant part of Theorem 3.3:

Theorem 3.4 (Theorem 1.4 from [12]). *Let T be a threshold circuit of depth 2 computing PARITY on 99% of all n-bit inputs. Then T has at least $\Omega(n)$ gates.*

Note this recovers the main theorem from the previous section for $d = 2$.

We now restate the random experiment that which we will use in this section.

Definition 1.1 (Blockwise Restriction). Let X be a variable set, and let P be a partitioning of X into $|P|/n$ consecutive blocks. Then define R_P as a distribution on assignments $\alpha : [n] \to \{0, 1, *\}$ that randomly fixes all but one element of each part of P.

We now state the main random restriction theorem to prove this lower bound.

Lemma 3.5 (Lemma 1.1 from [12]). *Let $\phi = (\mathbf{w}, \theta)$ be a threshold gate on n variables comprising the variable set X. Then*

$$\Pr_{\alpha \sim R_P}[f_\phi(\alpha) \text{ is not constant}] = O(|P|/\sqrt{n}).$$

We note some differences between this lemma and Lemma 2.2. One is that obviously the random procedure is much simpler and less targeted. Intuitively, we hope this allows us to achieve at least a more lucid upper bound on the failure probability. This will be necessary to show the required bound as it implies multiple assignments that can force the function to constant. We now sketch the proof of Theorem 3.3 using Lemma 3.5 as it requires a bit more thought than Theorem 2.1.

Proof of Theorem 3.3. Let T be a depth-2 Threshold Circuit on n bits with $N = o(\sqrt{n})$ gates that agrees with PARITY on at least 99% of the 2^n inputs. Let $|P| = 2$, that is R_P is the random restriction that generates a random assignment leaving exactly 2 inputs free. By Lemma 3.5 there are $O(N/\sqrt{n}) = o(1)$ non-trivial gates at the bottom level. By Markov's inequality, this means that there are *no* non-trivial gates on the bottom layer with at least 50% probability, i.e. on 2^{n-1} inputs, the remaining circuit on two inputs is equivalent to a single threshold gate. As previously stated, a threshold function is either increasing or decreasing in all its variables, meaning that there is at least $f(\alpha)$ can compute PARITY on at most 3 of the 4 possible inputs. Thus, it can only compute PARITY on at most $2^{n-1} * .75$ which is less than 99%. □

With our goal in mind, we now move on to explain the advancement achieved by the authors of [12].

3.1.1 Anti-Concentration

In the beginning of this section, we posited that what we would like to improve on is the specific targetting of favorable assignments we saw in Section 2.1. We observe that the proof of the last random restriction lemma made use of the fact that a gate can be eliminated simply by ensuring that "the right side" of its variables were set correctly. Somehow we would like to understand a different view of how a threshold gate might become constant under a restriction, one that is in a sense more fluid, numerically. To that end, we state without proof a lemma that will be fundamental to this goal.

Lemma 3.6 (Littlewood-Offord Lemma)**.** *Let $f(x) = \sum_{i=1} w_i x_i$ be a linear function and $r \in \mathbb{N}$. Let $I \subset \mathbb{R}$ be a finite interval and $|w_i| \geq |I|$ for at least r of the w_i. Then*

$$\Pr_{x \sim \{0,1\}^n}[f(x) \in I] \leq \frac{O(1)}{\sqrt{r}}.$$

We explain what makes this tool useful and also alude to how we will prove Lemma 3.5. For the sake of intuition, suppose our gate ϕ has weight and threshold $(\mathbf{w}, \theta)$ with variable set $[n]$, and suppose that $A, U \subset [n]$ respectively are the indices of the assigned and unassigned variables resulting from α. Then we observe that $f_\phi(\alpha)$ is constant iff $\sum_{i \in U} |w_i| > \theta - \sum_{i \in A} x_i w_i$, that is the sum of the assigned weights is greater than what the remaining weights can overcome. Although this is the same observation as is implicit in Lemma 2.4, there is more that can be said. The advancement comes from the fact that the combined sum of the assigned weights with the original threshold allows us to define a "bad" interval over the assigned weights where the gate does not become constant. Then, Lemma 3.6 can be used to bound the probability this happens. There is some technical work that is necessary to get the parameters in the correct form, but equiped with this intuition, we proceed with the formal proof.

Proof of Lemma 3.5. Let $\phi = (\mathbf{w}, \theta)$ be an threshold gate and R_P be the distribution as stated in the hypothesis of the lemma. Fix $A \subseteq [n]$ the indices of variables that will be assigned values and $U = [n] \setminus A$, the indices of variables kept unassigned. It will be convenient for the partition to be determined by U. Let $\alpha : [n] \to \{0, 1, *\}$ be a random assignment leaving the variables of U free. This allows us an equivalence to the statement we want to prove, that is

$$\Pr_{\alpha \sim R_P}\left[f_\phi(\alpha) \text{ is not constant}\right] = \mathbb{E}_U\left[\Pr_{\alpha \sim R_U}[f(\alpha) \text{ not constant}]\right]. \tag{3.1}$$

Let $f'(x) : \{0,1\}^{|U|} \to = \sum_{i \in A} w_i x_i$. Define an interval

$$I = \left[\theta - \sum_{i \in U, w_i > 0} |w_i|, \theta + \sum_{i \in U, w_i < 0} w_i \right]. \tag{3.2}$$

As previously stated if our assignment falls into this bad interval then our remaining function $f(\alpha)$ is not constant. Formally then we have that

$$\Pr_{\alpha \sim R_U} [f(\alpha) \text{ is constant}] = \Pr_{x \sim \{0,1\}^{|A|}} \left[f'(x) \in I \right].$$

We can remove the second condition from the summation from (3.2) by writing I as the union of intervals $I_i, |I_i| = |w_i|$. Thus, by union bound

$$\Pr_{\alpha \sim R_U} [f(\alpha) \text{ is constant}] \leq \sum_{i \in U} \Pr_{x \sim \{0,1\}^{|A|}} \left[f'(x) \in I_i \right].$$

We now want to choose a proper definition for r_i to use Lemma 3.6 for each of the summands. To that end, define r_i to be the rank of w_i if the weights were ordered greatest to least, i.e.:

$$r_i \geq r_j \longleftrightarrow |w_i| \leq |w_j|. \tag{3.3}$$

Lemma 3.7. $\Pr_{\alpha \sim R_U} [f(\alpha) \textit{ is not constant}] \leq \sum_{i \in U} \frac{O(1)}{\sqrt{r_i}}$

Note that this lemma does not come immediately from the Littlewood-Offord Lemma, as the intervals we have chosen are from the weights of A, while our current definition for r_i is the rank in the *complete* weights. Still, through some clever analysis we will prove that is the case. However, assuming the lemma, we complete the proof of Lemma 3.5. Recalling (3.1), we can

use Lemma 3.7 immediately to get that

$$\begin{aligned}
\Pr_{\alpha\sim R_P}\left[f_\phi(\alpha) \text{ is not constant}\right] &= \mathbb{E}_P\left[\Pr_{\alpha\sim R_U}\left[f(\alpha) \text{ not constant}\right]\right] && \text{(by (3.1))}\\
&\leq \mathbb{E}_U\left[\sum_{i\in U}\frac{O(1)}{\sqrt{r_i}}\right] && \text{(by Lemma 3.7)}\\
&= \sum_{i=1}^{n}\frac{O(1)}{\sqrt{r_i}}\cdot\Pr[i\in U] && \text{(by Linearity)}\\
&= \frac{|P|}{n}\cdot\sum_{i=1}^{n}\frac{O(1)}{\sqrt{r_i}} && \text{(by definition of } R_P)\\
&\leq \frac{|P|}{n}\cdot\sum_{l=1}^{n}\frac{O(1)}{\sqrt{l}} && \text{(reordering } r_i \text{ by rank)}\\
&= O(|P|/\sqrt{n}) && \text{(by } \sum_{i=1}^{n} i^{-1/2}\leq 2n^{1/2}),
\end{aligned}$$

which completes the analysis. It remains to show Lemma 3.7 and complete the proof.

Proof of Lemma 3.7. For our analysis, let r'_i be defined similarly to r_i, that is the rank of the ith weight, except only restricted to $i \in A$. We have that (3.3) holds the same for r'_i as it did for r_i. We also note a relation between r_i and r'_i along with the weights in U which occurs by definition of the assignment:

$$r_i \leq r'_i + \left[\ \#\text{ of } u\in U \text{ such that } |w_i|\leq|w_u|\right]. \tag{3.4}$$

Lemma 3.6 gives us that $\Pr_{x\sim\{0,1\}^{|A|}}\left[f'(x)\in I_i\right]\leq\frac{O(1)}{\sqrt{r'_i}}$. To show Lemma 3.7, it is equivalent to show that $r_i = O(r'_i)$. Obviously, $r'_i \leq r_i$. We show this by a case analysis.

1. The trivial case if if the weights that were left unassigned are packed together or representable by a single weight and rank. Formally, this would be if there is a $u\in U$ that we are summing over, such that at least $\frac{r_i}{2}$ other $v\in U$ have $r_v \geq r_u$. This allows us to immediately bound the sum from the claim by
$$\frac{r_i/2}{\sqrt{r_i}} > 1.$$

Thus the probability bound itself is trivial.

2. Assume that we are not in the trivial case then, i.e. that there is not one rank that can represent all the unassigned weights. or that for $u \in U$ there are at most $r_u/2$ different $v \in U$ with $r_u \leq r_v$. Recalling (3.4), this means that

$$r_i \leq r'_i + r_i/2 \leq 2k'_i.$$

This shows that $r_i = O(r'_i)$ as required, so the proof of Lemma 3.7 is complete and as stated previously, this proves Lemma 3.5. □

3.2 Modifications to prove wire lower bounds

We remark that the same observation we used to prove the wire lower bound of the previous chapter, can be used to modify the gate lower bound we just proved. We first restate what we would like to show.

Theorem 3.8. *Let T be a threshold circuit of depth 2 computing PARITY of all n-bit inputs. Then T has at least $\Omega(n^{3/2})$ wires.*

We write a similar technical lemma as Lemma 3.5.

Lemma 3.9 (Lemma 3.1 from [12])**.** *Let $\phi = (\boldsymbol{w}, \theta)$ be a threshold gate on n variables comprising the variable set X. Assume that ϕ depends only on $S \subset X$ of its inputs. Let P partition X into equal sizes and let R_P be the distributions on assignments $\alpha : [n] \to \{0,1,*\}$ that randomly fixes all but one element of each part of P. Then*

$$\Pr_{\alpha \sim R_P}[f_\phi(\alpha) \text{ is not a function of at most one input}] = O(|S|(|P|/n)^{3/2}).$$

Proof. We start in the same manner as in the proof of Lemma 3.5, which is to focus on a particular assignment U and take the expectation over it. Thus we have again that

$$\Pr_{\alpha\sim R_P}[f_\phi(\alpha) \text{ is a function of} \geq 1 \text{ input}] = \mathbb{E}_U\left[\Pr_{\alpha\sim R_U}[f_\phi(\alpha) \text{ is a function of} \geq 1 \text{ input}]\right]. \quad (3.5)$$

Like in the proof of Lemma 2.7, we can see that we've loosened the condition from Lemma 3.5 for which we consider a "success". This is because, as we argued, we can remove it from the level it either outputs exactly the input or its negation. Thus we can apply Lemma 3.7 to get that

$$\begin{aligned}\Pr_{\alpha\sim R_P}[f_\phi(\alpha) \text{ is not a function of at most one input}] &\leq \Pr_{\alpha\sim R_P}[f_\phi(\alpha) \text{ is not constant}] \\ &\leq \sum_{i\in U}\frac{O(1)}{\sqrt{r_i}}.\end{aligned}$$

We also observe that $f_\phi(\alpha)$ is a function on more than one input if and only if $|U\cap S|\geq 2$, which means we only need to consider U for which this is the case. Coupling this with the fact that $\sum_{i\in U\cap S}\frac{O(1)}{\sqrt{r_i}} \leq \sum_{i\in U}\frac{O(1)}{\sqrt{r_i}}$, we have that (3.5)

$$\begin{aligned}
&\leq \mathbb{E}_U\left[\Pr[|U\cap S|\geq 2]\dot{\sum}_{i\in U\cap S}\frac{O(1)}{\sqrt{r_i}}\right] \\
&= \sum_{i\in S}\frac{O(1)}{\sqrt{r_i}}\cdot\Pr[i\in U, |U\cap S|\geq 2] && \text{(U is now the only random variable)} \\
&\leq \sum_{i\in S}\frac{O(1)}{\sqrt{r_i}}\cdot\min\left\{\frac{|P|}{n}, \sum_{i\neq j\in S}\Pr[i,j\in U]\right\} && \text{(Prob. of both} \leq \text{min. of either)} \\
&= \sum_{i\in S}\frac{O(1)}{\sqrt{r_i}}\cdot\min\left\{\frac{|P|}{n}, |S|\cdot\frac{|P|^2}{n^2}\right\} \\
&\leq 2\sqrt{|S|}\cdot\min\left\{\frac{|P|}{n}, |S|\cdot\frac{|P|^2}{n^2}\right\} && \text{(by } \sum_{i\in S}\frac{1}{\sqrt{r_i}} \leq 2\sqrt{|S|}\text{)} \\
&\leq 2\sqrt{|S|}\cdot\left(|S|^{1/2}\cdot\frac{|P|^{3/2}}{n^{3/2}}\right) && \text{(}\min\{a,b\}\leq\sqrt{ab} \text{ if } a,b>0\text{)} \\
&= O\left(|S|\cdot\frac{|P|^{3/2}}{n^{3/2}}\right),
\end{aligned}$$

as required. □

3.3 Drawbacks and challenges

Thus, as it stands we have successfully shown tight lower bounds for the size of depth 2 threshold circuits approximating parity. However, again the methods used have their limitations. We observe that though the probability bounds given in this section were strong enough to give a large enough number of favorable assignments our circuit could not compute parity on, to do so required setting nearly every single variable. This essentially marks a death knell on extending this this random restriction lemma to higher depth in its current form, as we would not have enough variables left to set after just one iteration. All is not lost, however. In the following chapter, we will show how we again can refine our random restriction methodology to prove more general lower bounds.

Chapter 4

Simplification by Biasing

We saw previously how Kane and Williams extended the bounds of Impagliazzo et al to a new regime by making new observations on what it meant for a gate to become constant. In this chapter, we will show how we can take this one step further, to prove correlation bounds in the regime of arbitrary constant depth threshold circuits. The main advancement comes from a simple observation, which is that a gate becoming constant is not the only metric for simplification. Indeed, we saw a version of this notion when proving the wire bound in Chapter 2. Recall that our random restriction was not itself guaranteed to simplify all the gates to constants, rather we showed that some gates are left still with inputs to be simplified manually later.

Taking this into consideration, the advancement we will explain in this chapter is that we will not be expecting our gates to become constant following the proper paramertization of our random restriction. Instead, what we will expect and analyze is the probability that our gates become highly *biased*. We will argue that this is in a sense "good enough" for the bounds we would like to show, and leads to more favorable probability of success when analyzing the simplification of a gate or layer. We now state one of the main theorems of this section.

Theorem 4.1 (Corollary 3.1 from [4])**.** *Fix any $d \geq 2$. Assume that C is a depth-d threshold circuit over n variables with $k \leq n^{1/(2(d-1))}$ threshold gates and let $\delta = k/n^{1/(2(d-1))}$. Then, $Corr(C, Par_n) \leq O(\delta^{(1-1/d)})$. In particular, for any constant d, $Corr(C, Par_n) = o(1)$ unless $k = \Omega(n^{1/2(d-1)})$.*

To prove this result, we need to introduce new machinery to quantify and analyze bias.

4.1 Noise Sensitivity

Recall the random experiment from the previous section.

Definition 1.1 (Blockwise Restriction)**.** Let X be a variable set, and let P be a partitioning of X into $|P|/n$ consecutive blocks. Then define R_P as a distribution on assignments $\alpha : [n] \to \{0, 1, *\}$ that randomly fixes all but one element of each part of P.

We used such an experiment because we thought of $|P|$ as being a constant literal value, which made reasoning about correlation more straightforward. In this section it will be favorable to us in this section to now think of our assignments as being uniformly random subsets of size pn. Recall:

Definition 1.2 (Uniformly Random Subset)**.** We will use $\mathscr{R}_p^n$ to denote the distribution over restrictions, $\alpha : \{-1, 1, *\} \to \{-1, 1\}$, such that $\alpha(x) = *$ with probability p and $-1, 1$ with probability $(1-p)/2$ each.

We importantly note that the two distribution are identical for $pn = |P|$. This will allow us to use some of the results from the previous section here.

Definition 4.2. Let $f : \{-1, 1\} \to \{-, 1\}$ and $p \in [0, 1]$ a parameter. We define the *noise sensitivity of f with noise parameter p*, $\mathrm{NS}_p(f)$ as follows by experiment E. Given $x \in \{-1, 1\}^n$ selected uniformly at random, construct y from x by independently negating each bit of x with probability p. Then

$$\mathrm{NS}_p(f) = \Pr_{(x,y)\sim E}[f(x) \neq f(y)].$$

The random experiment in the aforementioned definition can also easily be expressed over the distribution of uniformly random subbsets.

Indeed, let x, y be generated equivalently as follows:

1. Let U be a randomly generated subset of $[n]$, such that $\Pr_{i\in[n]}[i \in U] = (1-2p)$. This is our set of surviving variables.

2. Uniformly at random, generate $s \sim \{-1,1\}^U$. Observe that $s \sim \mathscr{R}_p^n$.

3. Uniformly at random, generate $x', y' \sim \{-1,1\}^{[n]-U}$.

4. $\hat{x} = x' \parallel s, \hat{y} = y' \parallel s$.

One can verify then that $\Pr_{(\hat{x},\hat{y})}[f(\hat{x}) \neq f(\hat{y})] = \Pr_{(x,y)}[f(x) \neq f(y)]$, i.e. the distributions are equivalent, which implies the following fact.

Proposition 4.3. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be a boolean function. Then,*

1. *for $p \leq 1/2, NS_p(f) = (1/2)\mathbb{E}_{\alpha\sim\mathscr{R}_{2p}^n}[\mathrm{Var}(f(\alpha))]$.*

2. *for $p \geq 1/n$, $\mathrm{Corr}(f, PARITY_n) \leq O(NS_p(f))$.*

One can easily discern from this that a function's noise sensitivity is directly proportional to its correlation with the PARITY function. We can use this proposition and a theorem of Peres to actually show that a majority gate itself has low correlation with parity. This will serve as a basis for the intuition behind the proof and the induction itself.

Theorem 4.4 (Peres' Theorem, [14])**.** *Let $f : \{-1,1\}^n \to \{-1,1\}$ be any LTF. Then,*

$$\mathbb{E}_{\alpha\sim\mathscr{R}_p^n}[\mathrm{Var}(f(\alpha))] = NS_{p/2}(f) = O(\sqrt{p}).$$

Corollary 4.5. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be any threshold function. Then $Corr(f, Par_n) \leq O(n^{-1/2})$.*

Coupled with the above definitions, Peres' theorem essentially says that on application of a random restriction a threshold function becomes quite biased in expectation. Considering Fact 1.6, this means in fact that it can be well approximated by a constant function.

Remark 4.6. One might be tempted to try to compare the random restriction lemma from Kane and Williams with Peres' theorem. Indeed, spoken in terms of variance, Lemma 3.5 states that $\Pr_{\alpha\sim\mathscr{R}_n^p}[V>0] \leq O(p\sqrt{n})$, while Peres' theorem can be written to say that $\mathbb{E}_{\alpha\sim\mathscr{R}_n^p}[V] \leq \sqrt{p}$, where $V = \mathrm{Var}(f(\alpha))$. However, it is unclear on how to obtain an equivalency between the two, as a tool like Markov's inequality is undefined for parameter $k = 0$.

We will use these relations to prove Theorem 4.1. Like in previous sections, what we will prove in particular is the following depth reduction lemma.

Theorem 4.7. *Let C be a depth-d threshold circuit with at most k threshold gates. Then for any parameters $p,q \in [0,1]$, we have*

$$\mathrm{NS}_{p^{d-1}q}(C) \leq O(k\sqrt{p}+\sqrt{q}).$$

This theorem comes from a similar repeated application of a random restriction for each layer of a circuit as we saw in previous sections. Essentially, using our bound on the variance of the bottom layer gates, we replace all of them with their closest constant approximation. We repeat the procedure of random restriction $d-1$ times until the entire circuit becomes constant. This implies that the variance of the circuit must have been small, implying the noise sensitivity was small as well. Taken this way, we can say that Theorem 4.4 is the logical equivalent to a kind of random restriction lemma.

For completeness, we state how Theorem 4.1 can follow quite naturally with some simple parameterization.

Proof of Theorem 4.1. Apply Theorem 4.7 with

$$p = \frac{1}{n^{1/d}}\frac{1}{k^{2/d}}, q = \frac{k^{2/d}}{n^{1/d}}$$

so that $p^{d-1}q = 1/n$. This gives us

$$\begin{aligned}\mathrm{NS}_{1/n}(C) &\leq O\left(k\sqrt{\frac{1}{n^{1/d}}\frac{1}{k^{2/d}}} + \sqrt{\frac{k^{2/d}}{n^{1/d}}}\right) \\ &= O\left(\frac{k^{1-1/d}}{n^{1/(2d)}}\right).\end{aligned}$$

By Proposition 4.3, $\mathrm{Corr}(f, PARITY_n) \leq O\left(\frac{k^{1-1/d}}{n^{1/(2d)}}\right)$, which means that

$$\mathrm{Corr}(f, PARITY_n) \leq o(1) \text{ unless } k = \Omega(n^{1/2(d-1)}.$$

□

4.1.1 Noise Sensitivity Bounds via Peres' Theorem

The rest of this section is dedicated to proving Theorem 4.7. Again, using Proposition 4.3, we will instead show that

$$\mathbb{E}_{\alpha_d \sim \mathscr{R}^n_{p^d}}[\mathrm{Var}(C|_{\alpha_d})] \leq O\left(k\sqrt{p} + \sqrt{q}\right)$$

which will immediately imply Theorem 4.7. We will proceed as in previous sections, via an induction on d. As previously stated, the base case, $d = 1$ is simply Peres' Theorem.

For $d > 1$, let k_1 be the number of threshold gates at depth $d-1$. Let us analyze the result of a random restriction $\alpha \sim \mathscr{R}^n_p$ at the bottom level. Clearly we have that

$$\mathbb{E}_{\alpha_d}[\mathrm{Var}\left(C|_{\alpha_d}\right)] = \mathbb{E}_{\alpha}\left[\mathbb{E}_{\alpha_{d-1}}\left[\mathrm{Var}\left((C|_{\alpha})|_{\alpha_{d-1}}\right)\right]\right] \tag{4.1}$$

which will allow us to apply the induction hypothesis provided we can show that $C|_{\alpha}$ is of depth $d-1$. However, as previously stated, this is not actually necessarily true; the random restriction regime α does not immediately simplify $C|_{\alpha}$ to a depth $d-1$ circuit. Instead, what we will do is

give a new depth $d-1$ circuit C'_α and we will use the fact that the gates at the bottom are now highly biased to show that C'_α is a close approximation to $C|_\alpha$. We define C'_α as replacing all the gates $\phi|_\alpha$ at the bottom level of $C'|\alpha$ with their most probable constant $b_{\phi,\alpha}$. Recall that Peres' Theorem gives that

$$\mathbb{E}_\alpha\left[\mathrm{Var}(\phi|_\alpha)\right] \leq O(\sqrt{p}),$$

so using Fact 1.6 we can write that

$$\mathbb{E}_\alpha\left[\Pr_{x\in\{-1,1\}^{|\alpha^{-1}(*)}}[\phi|_\alpha(x) \neq b_{\phi,\alpha}]\right] \leq O(\sqrt{p}),$$

i.e. on average our replacement is inaccurate to the original gate a limited amount of times. Union bounding over the k_1 gates at the bottom gives that

$$\mathbb{E}_\alpha\left[\Pr_{x\in\{-1,1\}^{|\alpha^{-1}(*)}}[C|_\alpha(x) \neq C'_\alpha]\right] \leq k_1 O(\sqrt{p}). \tag{4.2}$$

The following elementary proposition allows us to write our desired quantity in terms of the expected variance of our new circuit.

Fact 4.8. *Let $f,g:\{-1,1\}^m \to \{-1,1\}$ and $\delta = \Pr_x[f(x)\neq g(x)$. Then for any $r\in[0,1]$, we have*

$$\mathbb{E}_{\alpha\sim\mathscr{R}^n_r}[\mathrm{Var}(f|_\alpha)] \leq \mathbb{E}_{\alpha\sim\mathscr{R}^n_r}[\mathrm{Var}(g|_\alpha)] + 4\delta.$$

Using Fact 4.8, we can rewrite (4.1) as follows:

$$\begin{aligned}
\mathbb{E}_{\alpha_d}[\mathrm{Var}\left(C|_{\alpha_d}\right)] &= \mathbb{E}_\alpha\left[\mathbb{E}_{\alpha_{d-1}}\left[\mathrm{Var}\left(\left(C|_\alpha\right)|_{\alpha_{d-1}}\right)\right]\right] && (4.1)\\
&= \mathbb{E}_\alpha\left[\mathbb{E}_{\alpha_{d-1}}\left[\mathrm{Var}\left(\left(C'_\alpha\right)|_{\alpha_{d-1}}\right)\right] + O(k_1\sqrt{p})\right] && \text{(Applying Fact 4.8 with (4.2))}\\
&= \mathbb{E}_\alpha\left[O(k-k_1)\sqrt{p}+\sqrt{q}+O(k_1\sqrt{p})\right]. && \text{(Induction Hypothesis)}
\end{aligned}$$

which simplifies to $O(k\sqrt{p}+\sqrt{q})$ as required, completing our induction, as well as the proof. □

4.2 A different approach

We have already seen that extending linear gate bounds to superlinear wire bounds seemed to require marginally closer and technical analysis. This remains true in this chapter. In fact, while we would hope to simply modify our noise sensativity result to achieve an understanding on wire complexity, that is unfortunately not the case. Recall that in our proof of Theorem 4.7 we union bounded over the bottom level gates to simplify the bottom level of the circuit. Unfortunately, this bound becomes too loose for any correlation bounds if the number of gates are superlinear, which they could be if the number of edges is superlinear. We thus dedicate this following section to explain a new approach to prove the correlation bounds when wire complexity is taken into account. We will show the following lower bound.

Theorem 4.9 (Theorem 4.2 from [4])**.** *For any $d \geq 1$, there is an $\varepsilon_d = (1/2)^{O(d)}$ such that any depth-d threshold circuit C with at most $n^{1+\varepsilon_d}$ wire that satisifies* $\mathrm{Corr}(C, PARITY_n \leq O(n^{-\varepsilon_d}))$.

4.2.1 How we might qualitatively refine noise sensitivity

As usual, we will first explain how we might try to extend the techniques or ideas of the previous lower bound to a gate bound. Like we saw in Section 2.1, now that we have a bound on the number of wires in our circuit, does this give us a metric by which the gates of our circuit can be ”close” to simplification? We already saw one version of this, that is that if we can reason about how close to constant the restricted function is, it may as well be constant in terms of its contribution to the complexity of the circuit.

We explain how this notion can actually be extended by referring back to Peres' theorem. Recall that Peres' theorem gave us that $\mathbb{E}_{\alpha \sim \mathscr{R}_n^p}[V] \leq \sqrt{p}$, where $V = \mathrm{Var}(f(\alpha))$. Applying Markov's rule arbitrarily with paramter k, this says that $1/k$ of the functions hit with the random restriction, will have variance at most $O(k\sqrt{p})$. What does this mean? Well, now naively

applying Chebyshev's inequality with parameter a, we could get for such a function f that

$$\Pr_{x\in\{-1,1\}^{Supp(f)}}[|f(x)-\mathbb{E}(f(x))|\geq a]\leq\frac{k\sqrt{p}}{a^2}.$$

And therein lies our advancement. Namely, if we could obtain some quantification of the expected value of a threshold gate, we may be able to obtain additional savings on some of the gates at the bottom layer. Doing this in the previous regime, where we were concerned with the number of gates, may have been more difficult because a $1/k$ failure rate may have been hard to overcome. But, as we saw in Section 2.2, we can exploit the limit on the budget of wires to show that we can manually simplify the problem gates by setting a small number of variables.

In the following subsection, we will make plain this intuition, as well as state the random restriction lemma that we will use in the proof of Theorem 4.9.

4.2.2 Balance: a new notion of bias

Although we just described how we might use Chebyshev's Inequality and the expected value of the threshold function to refine the proof of Theorem 4.7, there is actually a better way, one that does away with the somewhat bothersome fact that a threshold function has a signed output.

Recall that idea behind using Lemma 3.6 was to bound the probability our gates became constant under the author's random restriction. They became constant because the random restriction was likely to set the variables of the function in such a way that the linear function on the remaining variables had no inputs that could exceed (or precede) the threshold. In other words, the threshold of the resulting function θ' was greater in *magnitude* than the maximum value of the linear function on the remaining weights (w'). This intuition can be generalized, i.e. if the threshold value is much larger than the average sum of the weights of the threshold gate, it will intrinsically be highly biased.

Lemma 4.10 (Modified Chernoff Bound)**.** *Let $w\in\mathbb{R}^n$ be arbitrary and x is chosen uniformly*

from $\{-1,1\}^n$. *Then*

$$\Pr_x[|\langle w,x\rangle| \geq t * \|w\|_2] \leq \exp(-\Omega(t^2)).$$

This immediately leads to our new notion of bias.

Definition 4.11 (t-Imbalance)**.** We say that a threshold gate $f = (w,\theta)$ is t-imbalanced if $|\theta| \geq t * \|w\|_2$ and t-balanced if $|\theta| < t * \|w\|_2$.

Now we have all we need to state the author's random restriction lemma.

Lemma 4.12 (Lemma 4.4 in [4])**.** *The following holds for some absolute constant* $p_0 \in [0,1]$. *For any threshold gate* f *over* n *variables with label* (w,θ) *and any* $p \in [0,p_0]$, *we have*

$$\Pr_{\alpha\sim\mathscr{R}_p^n}\left[f(\alpha) \textit{ is not } \frac{1}{p^{\Omega(1)}}\textit{-balanced}\right] \leq p^{\Omega(1)}.$$

We defer the proof of this lemma to a later section. For now, we would like to prove the last main theorem of this thesis.

4.2.3 Using balance to achieve correlation bounds

The proof of Theorem 4.9 will be an induction on the depth of d. The base case $d = 1$ of a single majority gate is already proven via Peres' theorem (see Corollary 4.5). For the induction step, $d > 1$, first define parameters in terms of a large enough constant B:

$$\varepsilon = B^{-2d-1} \quad \delta_d = B\varepsilon_d \quad p = n^{-\delta_d/2}$$
$$t = p^{-1/3} \qquad q = \tfrac{1}{t} \qquad D = \tfrac{1}{6}.$$

Let C be a circuit on n variables with $n^{1+\varepsilon_d}$ wires. We want to calculate $\mathrm{Corr}(C, PARITY_n)$ in terms of a lower depth circuit C_{d-1} on n_{d-1} variables such that

$$n^{1+\varepsilon_d} \leq n_{d-1}^{1+\varepsilon_{d-1}}.$$

We will do this in multiple steps. First, consider the random assignment $\alpha = (A,U,w) \sim \mathscr{R}_p^n$. Like in the proofs of Theorems 3.9 and 4.1, if applied to the circuit C this random restriction will

intrinsically trivialize some gates so we can remove/replace them in the circuit without losing too much correlation with the original circuit. The gates that are not easily simplified will have to be simplified manually by setting all the wires into these gates like in Section 2.2.

We first argue that we will not need to set many variables for these problem gates in expectation. To that end, we divide the gates at the bottom level into two classes, small and large. These, we will analyze separately, counting how many wires we need to set for each class. We will show that the small gates are likely to already become constants or single-wired gates (except for the ones we will have to simplify manually) while the large gates are likely to become biased. Let S, L be the indices of such gates where

$$i \in \begin{cases} S \text{ if } \mathrm{FanIn}(f_i) \leq p^{-D} \\ L \text{ if } \mathrm{FanIn}(f_i) > p^{-D} \end{cases}.$$

where D is a parameter we will optimize later.

We will also assume that the variable sets of our restriction (A,U) can be considered as somewhat "generic". This will allow us to ensure we have favorable bounds on both the total number of unassigned variables we start with *and* the possible number of variables we will need to set later.

Claim 4.13. Let $\alpha = (A,U,y)$ be as was just defined. Call (A,U) a generic restriction set if the following is true for α:

1. Let ϕ_i be a large gate in the bottom level such that $i \in L$. Then $\mathrm{FanIn}(\phi|_\alpha)) \leq 2p\mathrm{FanIn}(\phi)$.
2. $|U| \geq \frac{np}{2}$.

Let $\mathscr{G}_\alpha$ be the event that a restriction is generic. Then

$$\Pr_\alpha[\neg\mathscr{G}] \leq \exp(-n^{\Omega(\delta_d)}).$$

Proof of Claim 4.13. We want to analyze the probability that a given restriction set (A,U) is not generic. We have two criteria we need to check:

1. $i \in L \to \text{FanIn}(f_i|_\alpha) \leq 2p * \text{FanIn}(f_i)$. Since each variable is set to some constant with probability $1-p$, $\mathbb{E}[\text{FanIn}(f_i|_\alpha)] = p \cdot \text{FanIn}(f_i)$. The expected fan-in of each gate is at least $p^{-D} = n^{D\delta/2}$. By applying a Chernoff bound (Theorem 1.3), we have

$$\Pr[\text{FanIn}(f_i|_\alpha) > 2p \cdot \text{FanIn}(f_i)] \leq \exp(-\Omega(4n^{D\delta/2})).$$

2. $|U| \geq \frac{np}{2}$. We similarly apply another Chernoff bound to get that

$$\Pr[|U| < \frac{np}{2}] \leq \exp(\Omega(np)).$$

We combine the above quantities, using a union bound over the large gates for item 1. (Note that there can be at most $\frac{n^{1+\varepsilon_d}}{p^{-D}} = n^{1+\varepsilon_d - D\delta_d/2} = \ell$ large gates) to get

$$\Pr_\alpha[\neg\mathscr{G}] \leq \ell \exp(\Omega(-n^{D\delta_d/2})) + \exp(\Omega(np)) \leq \exp(-n^{D\delta_d/4}) \leq n^{\Omega(\delta_d)})$$

for large enough B. □

Since there are many generic restrictions, assuming our random restriction is on a generic set does little to the success probabilities of our random restriction lemmas (Lemmas 4.12 and 3.9). This leads to the following claim.

Claim 4.14. For $i \in L$,

$$\Pr_\alpha[f_i(\alpha) \text{ is not } q\text{-imbalanced } |\mathscr{G}] \leq 2q,$$

and for $i \in S$

$$\Pr_\alpha[f_i(\alpha) \text{ is a function of } > 1 \text{ input } |\mathscr{G}] \leq 2p^{3/2}p^{-D}.$$

Proof of Claim 4.14. We leave the reader to verify this using Claim 4.13. □

This will allow us to bound the number of variables we need to set for the problem gates. We write for $i \in S$

$$Y_i^S = \begin{cases} 0 \text{ if } f_i(\alpha) \text{ is not constant or single wired} \\ \text{FanIn}(f_i(\alpha)) \text{ otherwise} \end{cases},$$

and for $i \in L$,

$$Y_i^L = \begin{cases} 0 \text{ if } f_i \text{ is not q-imbalanced} \\ \text{FanIn}(f_i(\alpha)) \text{ otherwise} \end{cases}.$$

Thus the number of wires we need to set for each group is $Y^S = \sum_i Y_i^S, Y^L = \sum_i Y_i^L$ and $Y = Y^S + Y^L$.

Claim 4.15. Let Y be as above. Then,

$$\mathbb{E}[Y \mid \mathscr{G}] \leq 4p^{4/3} \cdot (n^{1+\varepsilon_d}).$$

Proof of Claim 4.15. As alluded to, we will consider small and large gates separately.

1. Large gates, i.e. $f_i \in L$.

   Recall that since we conditioned on our restriction being generic $\text{FanIn}(f_i(\alpha))$ could be at most $2p \cdot \text{FanIn}(f_i(\alpha))$. Thus

   $$\mathbb{E}[Y_i|\mathscr{G}] \leq (2p * \text{FanIn}(f_i(\alpha)) * \Pr_\alpha[f_i(\alpha) \text{ is not } q\text{-imbalanced } |\mathscr{G}] \leq 4pq * \text{FanIn}(f_i(\alpha))$$

   where the second inequality follows from Claim 4.14. Then using linearity we have that

   $$\mathbb{E}[Y^L \mid \mathscr{G}] \leq 4pq \sum_{i \in L} \text{FanIn}(f_i(\alpha)).$$

2. Small gates, i.e. $f_i \in S$.

Again using Claim 4.14 and linearity, we have

$$\mathbb{E}[Y^S \mid \mathscr{G}] \leq 2p^{3/2}p^{-D}\sum_{i\in S}\text{FanIn}(f_i(\alpha)).$$

Taking the two together, we have that

$$\begin{aligned}\mathbb{E}[Y|\mathscr{G}] &\leq 4pq\cdot\sum_{i\in L}\text{FanIn}(f_i(\alpha)) + p^{3/2}p^{-D}\sum_{i\in S}fanin(f_i(\alpha)) \\ &\leq \max\left\{4pq, 2p^{3/2}p^{-D}\right\}\cdot n^{1+\varepsilon_d} \leq 4p^{4/3}\cdot(n^{1+\varepsilon_d})\end{aligned}$$

because we have that $4pq = 4p^{4/3}, 2p^{3/2}p^{1/6} = 2p^{4/3}$. □

We want specifically to condition on the expectation from Claim 4.15 being achieved (within a q factor), so that we can simplify the bottom level as we have stated above. Using Markov's inequality (Theorem 1.4), we can argue that the fraction of inputs on which this does not occur is small enough not to affect our result.

Claim 4.16. Fix a generic set (A', U'). Call $\alpha_1 = (A', U', y)$ a "good" restriction if $f_C(\alpha)$ is such that the total fan-in into large t-balanced gates or small non-constant/single-wire gates is at most $\mu/\sqrt{q}$. Then

$$\Pr_{y\in\{-1,1\}^A}[\alpha_1 \text{ is not good}] \leq 2\sqrt{q}.$$

This ensures us that we will be able to take care of all the problem gates. We now move onto simplifying the circuits on the so-called "good" restrictions. Using Lemma 4.12, we have that many of them will become highly biased, allowing us to set them to constants as we did in the proof of Theorem 4.1. We will use the following claim:

Claim 4.17. Let $C|_{\alpha_1}$ be the circuit restricted to any good restriction α_1. Then there is a circuit C' which replaces all the t-imbalanced gates of C_{α_1} with their most probable constant such that

$$\Pr_x[C|_{\alpha_1}(x) \neq C'(x)] \leq \exp(n^{-\varepsilon_d}).$$

Proof of Claim 4.17. What we will do is set all t-imbalanced gates to their most probable constants. Formally, consider a t-imbalanced threshold gate $f = (w, \theta)$. By definition $|\theta| \geq t \cdot \|w\|_2$. Replace the t-imbalanced gates f_i with constant gates b_i such that

$$b_i = \begin{cases} 1 \text{ if } \theta \geq t \cdot \|w\|_2 \\ -1 \text{ if } -\theta \geq t \cdot \|w\|_2 \end{cases}.$$

This transforms $C|_{\alpha_1}$ into C'. Note that $C|_{\alpha_1}(x) \neq C'(x)$ iff there is a t-imbalanced threshold gate such that $f_i(x) \neq b_i(x)$. By the Chernoff bound (Theorem 1.3), we have

$$\Pr_{x \in \{-1,1\}}[f_i(x) \neq b_i(x)] \leq \exp(-\Omega(t^2)) \leq \exp(-n^{\Omega(\delta_d)}).$$

Union bounding over the maximum $n^{1+\varepsilon_d}$ gates

$$\Pr_{x \in \mathbb{R}^n}[C|_{\alpha_1}(x) \neq C'(x)] \leq n^{1+\varepsilon_d} \cdot \exp(-n^{\Omega(\delta_d)}) \leq \exp(-n^{\varepsilon_d}).$$

□

All that is left is to then simplify C', which is now guaranteed to be only small gates and large t-balanced gates. As this is the last step, we show how to simplify C' and prove the main theorem all together.

Completion of Theorem 4.9. Recall that we wanted to calculate $Corr(C, Parity_n)$ in terms of a low depth circuit C_{d-1}. First, observe that the generic set from Claim 4.13 defines a restriction

tree $T_{A'}$ such that by Fact 1.7, we can write

$$\begin{aligned}
\mathrm{Corr}(C, PARITY_n) &\leq \mathbb{E}_{\ell \sim T_{A'}}[\mathrm{Corr}(C|_{\alpha_\ell}, PARITY_{|U'|})] \\
&\leq 2\sqrt{q} + \max_{\alpha_\ell \text{ good}}[\mathrm{Corr}(C|_{\alpha_\ell}, PARITY_{|U'|})] && \text{(By Claim 4.13)} \\
&\leq 2\sqrt{q} + \mathrm{Corr}(C', PARITY_{|U'|}) + \exp(n^{-\varepsilon_d}) && \text{(Claim 4.17)} \\
&\leq 2n^{-\varepsilon_d} + \mathrm{Corr}(C', PARITY_{|U'|})
\end{aligned}$$

where C' has no t-imbalanced gates. Since α_ℓ was a good restriction, we also have total fan-in into the problem gates of C' is at most $\mu/\sqrt{q}$. Thus, we can generate another restriction tree T_2 that is *not* random, that sets the the inputs to all the large t-balanced gates and all the small non-constant/not-single-wire-gates. Applying Fact 1.7 again, we have

$$\mathrm{Corr}(C, PARITY_n) \leq 2n^{-\varepsilon_d} + \mathbb{E}_{\ell_2 \sim T_2}[\mathrm{Corr}(C'|_{\ell_2}, PARITY_{|U'|-\frac{\mu}{\sqrt{q}}})]. \tag{4.3}$$

However, by definition, all $C'|_{\ell_2}$ are depth $d-1$.

How many variables survived in our series of restrictions? At least $pn/2$ survived the original restriction, and to clean up the remaining problem gates we set at most $\mu/\sqrt{q}$ more.

$$\begin{aligned}
\mu/\sqrt{q} &= 4p^{7/6} * n^{+\varepsilon_d} \\
\mu/\sqrt{q} &\leq 4np * n^{\varepsilon_d - B\varepsilon_d}
\end{aligned}$$

which is at most half the variables we set in the first restriction, for B large enough. This means at least $n_1 = pn/4$ variables survived in total. How many wires does the circuit have? We would like it to have $n_1^{1+\varepsilon_{d-1}}$ for

$$\varepsilon_{d-1} = B^{-2(d-1)-1}$$

so that we can use our induction hypothesis. We leave that to the reader to verify. Thus, we have

that

$$\mathrm{Corr}(C'|_{\ell_2}, PARITY_{|U'|-\frac{\mu}{\sqrt{q}}}) \leq n^{\varepsilon_{d-1}}$$

by our induction hypothesis and plugging this into (4.3) completes the induction. □

4.2.4 Proving the main restriction lemma

In this section, we give the promised proof of the main random restriction lemma of this section. We state it again as a reminder.

Lemma 4.12 (Lemma 4.4 in [4])**.** *The following holds for some absolute constant $p_0 \in [0,1]$. For any threshold gate f over n variables with label (w,θ) and any $p \in [0,p_0]$, we have*

$$\Pr_{\alpha \sim \mathscr{R}_p^n}\left[f(\alpha) \text{ is not } \frac{1}{p^{\Omega(1)}}\text{-balanced}\right] \leq p^{\Omega(1)}.$$

As was the case in Chapter 3, the quantity we are concerned with $\|w'\|_2$ is determined by the result of a random sum, $w' = \Theta - \sum_i w_i x_i$. Thus, we can utilize a similar anti-concentration bound, one that is a special derivative of the Berry-Essen Theorem.

Proposition 4.18 (Anticoncentration for regular linear functions)**.** *Let $x_1,\ldots,x_n$ denote independent uniformly ± 1 and let $w_1,\ldots,w_n \in \mathbb{R}$. Write $\sigma = \sqrt{\sum_i w_i^2}$, and assume that $|w_i|/\sigma \leq \tau$ for all i. Then for any interval $[a,b] \subset \mathbb{R}$,*

$$|\Pr[a \leq w_i x_i + \cdot + w_n x_n \leq b] - \Psi\left(\left[\frac{a}{\sigma},\frac{b}{\sigma}\right]\right)| \leq 2\tau,$$

where $\Psi([c,d]) \coloneqq \Psi(d) - \Psi(c)$. In particular,

$$|\Pr[a \leq w_i x_i + \cdot + w_n x_n \leq b] \leq \frac{|b-a|}{\sigma} + 2\tau,$$

Observe that this result presumes a certain "niceness" of the gate we are analyzing, which is that its weights are, in a sense, regularly spread ($|w_i|/\|w\|_2 \leq 2\tau$). Functions of this kind can

be thought of as being essentially the majority function, which proves to be the easiest case to prove for Lemma 4.12.

To that end, we first give a proof of Lemma 4.12, assuming the gate is the majority gate.

Proof of Lemma 4.12: Majority. Denote $t = \frac{1}{p^{\Omega(1)}}$. Majority is the easy case because in reference to the parameters of Proposition 4.18 are already known: $\sigma = \sqrt{n}, \tau = \frac{1}{\sqrt{n}}$ for an n-input MAJORITY function.

Let $f = (\theta, w)$ be an n bit Majority gate. Consider a random restriction α where $A, U \subset [n]$ are the indices of the assigned and unassigned variables. This time let $\Pr[x \in U] = p$ with uniform probability, and let assigned variables be assigned values ± 1 with equal probability. Let $f(\alpha) = (\theta', w'), \theta' = \theta - \sum_{i \in A} x_i w_i$, and w' be the weights of the surviving variables. For ease of notation relating to Proposition 4.18, we will write $\sum_{i \in A} x_i w_i$ as $\langle y, w'' \rangle$ where $y = \alpha(A)$ and w'' are the weights associated with the assigned variables. In essence what we want to show is to bound the probability that $\theta < t\|w'\|_2$.

We first observe that $\mathbb{E}_\alpha[\|w'\|_2^2] = p * \|w\|_2^2 = pn$. Since the probability that a given variable (and its corresponding weight) survives is independent of the others, we can use Chernoff bounds to get

$$\Pr[\|w'\|_2^2 \geq 2p * \|w\|_2^2] \leq \Pr[\|w'\|_2^2 \geq 2pn] \leq \exp(-n/4). \tag{4.4}$$

Thus, we have $\Pr[f(\alpha) \text{ is not } t\text{-balanced}] \leq$

$$\Pr\left[f(\alpha) \text{ is not } t\text{-balanced} \mid \|w'\|_2^2 \geq 2p\|w'\|_2^2\right] + 1 \cdot \exp(-n/4),$$

so for the rest of the proof we will assume assume that $\|w'\|_2^2 > 2p * \|w\|_2^2$. This allows us to

write

$$\begin{aligned}
\|w'\|_2^2 &< p * \|w\|_2^2 \\
\|w'\|_2 &< \sqrt{p * \|w\|_2^2} \\
\|w\|_2 - \|w''\|_2 &< \sqrt{p * \|w\|_2^2} \\
&= \sqrt{2p}\|w\|_2 \\
\|w\|\sqrt{1-2p} &< \|w''\|_2 \\
\sqrt{n/2} &\leq \|w''\|_2. \qquad \text{(Choose } p < p_0 \text{ accordingly)}
\end{aligned}$$

Thus since $\|w''\|_2 \geq \|w\|/2$, we have by our earlier assumption that

$$\|w'\|_2 \leq 4\sqrt{p}\|w'\|. \tag{4.5}$$

Recall that we had by definition the relation

$$\begin{aligned}
|\theta'| &\geq t\|w'\|_2 \\
|\theta - \langle w'', y\rangle| &\geq t\|w'\|_2 \\
&\geq t \cdot 4\sqrt{p}\|w''\|_2. \qquad \text{(By (4.5))}
\end{aligned}$$

Our final step is to use our anti-concentration theorem to bound the probability that this occurs. Again, we want to bound $\Pr[|\theta - \langle w'', y\rangle| \geq t\|w'\|_2]$, which we write as the probability that

$\langle w'', y\rangle$ lies in a "bad" interval. $\Pr[|\theta - \langle w'', y\rangle| \geq t\|w'\|_2]$

$$
\begin{aligned}
&= \Pr[\langle w'', y\rangle \in [\theta - 4t\sqrt{p} * \|w''\|_2, \theta + 4t\sqrt{p} * \|w''\|_2]] && \text{(By definition)} \\
&\leq \frac{4t\sqrt{p}\|w''\|_2}{\|w''\|_2} + 2\tau && \text{(By Proposition 4.18)} \qquad (4.6) \\
&\leq 4t\sqrt{p} + \frac{1}{2\sqrt{n}}.
\end{aligned}
$$

Since $\frac{1}{2\sqrt{n}} \to 0$, this completes our proof because we thought of $t = \frac{1}{p^{\Omega(1)}}$. □

Biasing general Linear Threshold Functions

We now continue to prove Lemma 4.12 for general threshold functions. The main idea behind the proof is to them into one of three kinds:

1. Gates where the weights are evenly spread, such that none of their variables have too much weight (called *regular*; these can be thought of as being similar to MAJORITY)

2. Gates that are top heavy, and have some *critical* variables with large weight. We then sub-divide these gates into

   (a) Gates where the large weight is concentrated in only a few variables.

   (b) Gates where the large weight is concentrated in relatively many.

We now formalize regularity and criticality.

Definition 4.19. Let $\varepsilon \in [0,1]$ be a real parameter. We say that $w \in \mathbb{R}^n$ is *ε-regular* if for each $i \in [n]$, $|w_i| \leq \varepsilon * \|w\|_2$. Assume now that the coordinates of the vector are sorted by decreasing magnitude ($|w_1| \leq |w_2| \leq \cdots \leq |w_n|$), then the *$\varepsilon$-critical index of w*, $K = K(\varepsilon)$ is the least index such that $w_{\geq K+1}$ is ε-regular. Sometimes we will refer to the variables removed as *critical variables*. $K = 0$ if w is already ε-regular and $K = n$ if there is no such index that makes $w_{\geq K+1}$ ε-regular. Accordingly, a gate $f = (w, \theta)$ is ε-regular if w is, and it's ε-critical index is defined to be the ε-critical index of w.

We now sketch the proof of Lemma 4.12 for each of these cases. We define

$$\varepsilon' \leq p^{1/8},\ L = L(\varepsilon') = \frac{100\log^2(1/\varepsilon')}{(\varepsilon')^2}$$

to differentiate the cases. Assume that $p \leq p_0$ such that $p^{1/8} \leq \frac{1}{\sqrt{16\log(1/p)}}$.

1. Case 1: ε'-regular functions.

   Proof. One can refer simply to the proof from the previous section, replacing the values for the weights in (4.4) and (4.6)) with $\|w\|_2$ and ε' respectively. We leave it up to the reader to verify the calculations. □

2. Case 2a: f is NOT ε'-regular and has critical index $K \leq L$ (bounded number of critical variables).

   Proof. What we would like to show is that since there are relatively few critical variables, it is very likely that we set all of them, meaning that the resulting function is itself regular (we can appeal to the regular case). To that end, let

   $$\mathrm{B}^{c,\ell}(\alpha) \text{ be the event that } U \text{ contains at least } c \text{ of the } \ell \text{ most critical variables.}$$

   Then, by definition

   $$\Pr_{\alpha}[\mathrm{B}^{k,\ell}(\alpha)] \leq (ep\ell/k)^k \tag{4.7}$$

   because there are $\binom{\ell}{k} \leq (e\ell/k)^k$ many subsets of the critical variables, each with p^k chance of remaining unassigned in U. The event that we leave any critical variable free $\mathrm{B}^{1,K}(\alpha)$ then has probability $epK \leq epL \leq \sqrt{p}$ of occurring so we have

   $$\begin{aligned}\Pr\left[f|_\alpha \text{ is not } t\text{-imbalanced}\right] &\leq \Pr[\mathrm{B}^{1,K}] + \Pr\left[f|_\alpha \text{ is not } t\text{-imbalanced}|\neg\mathrm{B}^{1,K}\right] \\ &\leq \sqrt{p} + \Pr\left[f|_\alpha \text{ is not } t\text{-imbalanced}|\neg\mathrm{B}^{1,K}\right].\end{aligned} \tag{4.8}$$

We can think of all the critical variables as being set "first", since they are done independently, so our analysis of the previous case can be applied to the restriction alpha giving us

$$\Pr[f|_\alpha \text{ is not } t\text{-imbalanced}] \leq \sqrt{p} + p^{\Omega(1)} \leq p^{\Omega(1)}.$$

□

3. Case 2b: f is NOT ε'-regular and has critical index $K > L$ (possibly many critical variables).

Proof. We start again in the same vein as (4.8):

$$\begin{aligned} \Pr[f|_\alpha \text{ is not } t\text{-imbalanced}] &\leq \Pr[\mathrm{B}^{1,L}] + \Pr\left[f|_\alpha \text{ is not } t\text{-imbalanced}|\neg \mathrm{B}^{1,L}\right] \\ &\leq \sqrt{p} + \Pr\left[f|_\alpha \text{ is not } t\text{-imbalanced}|\neg \mathrm{B}^{1,L}\right]. \end{aligned} \tag{4.9}$$

Again, we condition on a fixed I so that $\mathrm{B}^{1,L}$ does not occur. Let $L_0 = (1/(\varepsilon')^2) \cdot 3log(1/\varepsilon')$.

Claim 4.20. Let $i < L - L_0$. Then:

$$\|w_{\geq(i+L_0)}\|_2^2 \leq \frac{(\varepsilon')^2}{9} \cdot \|w_{\geq i}\|^2 \leq \frac{w_i^2}{9}.$$

We leave the proof of this claim to the reader. We can get from it the following Corollary.

Corollary 4.21. *Let* $i_1 = 1, i_2 = 1 + L_0, \cdots, i_{r+2} = 1 + (r+1)L_0 \leq L'$. *Then*

$$w_{i_{j+1}} \leq \frac{w_{i_j}}{3} \qquad \textit{and} \qquad \|w_{\geq i_{j+1}}\|_2^2 \leq \frac{(\varepsilon')^2}{9} \cdot \|w_{\geq i_j}\|_2^2$$

where r is maximized by $10log(1/\varepsilon')$.

Essentially what this says is that there is a geometric drop off of the weights and their

volume. Thus, we can bound the weights of the unassigned variables, because we know that at least the top L heaviest weights were not in the unassigned set. Again, let A, U be the assigned and unassigned variables. Then

$$\begin{aligned}
\sum_{i\in U} w_i^2 &\leq \|w_{\leq L}\|_2^2 \\
&\leq \|w_{\geq L}\|_2^2 \\
&\leq \|w_{\geq i_{r+2}}\|_2^2 && \text{(By Corollary 4.21)} \\
&\leq \frac{(\varepsilon')^2}{9}\|w_{\geq i_{r+1}}\|_2^2 \\
&\leq \frac{(\varepsilon')^2}{81} w_{i_r}^2. && \text{(By Claim 4.20)} \qquad (4.10)
\end{aligned}$$

Recall that our original goal was to understand how the threshold of the restricted function θ' related to the norm of the remaining weights. In the case of the regular function/Majority, we wrote $\theta' = \theta - \sum_{1\in A} x_i w_i$. we condition on setting all the variables in A other than $x_{i_1}, ..., x_{i_r}$ which we know also must have been assigned and we let $\Theta = \theta - \sum_{j\notin[r]} w_{i_j} y_{i_j} \in \mathbb{R}$. Then

$$\theta' = \Theta - \sum_{j\in[r]} w_{i_j} y_{i_j}. \qquad (4.11)$$

The probability that the function is not t-imbalanced is at most

$$\Pr_{x_{i_1},...,x_{i_r}}\left[|\theta'| \leq \frac{1}{p^{16}}\sqrt{\sum_{i\in U} w_i^2}\right] \leq \Pr_{x_{i_1},...,x_{i_r}}\left[|\theta'| \leq \frac{1}{9}|w_{i_r}|\right] \qquad \text{(By (4.10), } \varepsilon' = p^{1/8}).$$

Finally, we plug in for (4.11) to get that this probability is at most:

$$\Pr_{x_{i_1},...,x_{i_r}}\left[\sum_{j=1}^{r} w_{i_j} x_{i_j} \in [\Theta - \frac{1}{9}|w_{i_r}|, \Theta + \frac{1}{9}|w_{i_r}|]\right].$$

One can verify that there can be at most one choice of $x_{i_1}, ..., y_{i_r}$ such that $\sum_j w_{i_j} y_{i_j}$ lies in the required interval (see [5], Claim 5.7). Thus, $\Pr\left[f|_\alpha \text{ is not } t\text{-imbalanced}|\neg \mathrm{B}^{1,L}\right] 2^{-r} \leq$

$(\varepsilon')^{10} \leq p$ meaning $\Pr[f|_\alpha$ is not t-imbalanced$] \leq \sqrt{p} + p$ as required. □

4.3 Conclusion and next steps

In this chapter, we showcased the latest results concerning size-depth tradeoffs for constant depth threshold circuits. However, in this thesis, we explained how such results came to be, as a progression and refinement of a limited number of tools. This speaks especially to the power of random restriction techniques in proving circuit lower bounds in general.

However, as mentioned by all the main authors discussed in this thesis, in this regime such techniques seem to be nearing the limits of what is possible given known upper bounds. Despite this there are still many questions one could pose following the results explained in this chapter. For example, is the refinement (that is shifting our focus to gate constancy to gate bias) given in this chapter applicable in other situations? Does it, in fact, give a quantitative or qualitative improvement over the standard technique? In the same vein, are there other approaches to gate simplification that have not been considered? In the future, we would like to look more into answering these questions and others.

Bibliography

[1] Miklós Ajtai. 11-formulae on finite structures. *Annals of pure and applied logic*, 24(1):1–48, 1983.

[2] Ravi B Boppana and Michael Sipser. The complexity of finite functions. In *Algorithms and complexity*, pages 757–804. Elsevier, 1990.

[3] Marco Carmosino, Kenneth Hoover, Russell Impagliazzo, Valentine Kabanets, and Antonina Kolokolova. Lifting for Constant-Depth Circuits and Applications to MCSP. In Nikhil Bansal, Emanuela Merelli, and James Worrell, editors, *48th International Colloquium on Automata, Languages, and Programming (ICALP 2021)*, volume 198 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 44:1–44:20, Dagstuhl, Germany, 2021. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[4] Ruiwen Chen, Rahul Santhanam, and Srikanth Srinivasan. Average-case lower bounds and satisfiability algorithms for small threshold circuits. *CoRR*, abs/1806.06290, 2018.

[5] Ilias Diakonikolas, Parikshit Gopalan, Ragesh Jaiswal, Rocco Servedio, and Emanuele Viola. Bounded independence fools halfspaces, 2009.

[6] Paul E Dunne. *The complexity of Boolean networks*. Academic Press Professional, Inc., 1988.

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

[8] John Hastad. Almost optimal lower bounds for small depth circuits. In *Proceedings of the eighteenth annual ACM symposium on Theory of computing*, pages 6–20, 1986.

[9] William Hesse. Division is in uniform tc0. In *International Colloquium on Automata, Languages, and Programming*, pages 104–114. Springer, 2001.

[10] John E Hopcroft, Rajeev Motwani, and Jeffrey D Ullman. Introduction to automata theory, languages, and computation. *Acm Sigact News*, 32(1):60–65, 2001.

[11] Russell Impagliazzo, Ramamohan Paturi, and Michael E. Saks. Size-depth trade-offs for threshold circuits. In *Proceedings of the Twenty-Fifth Annual ACM Symposium on Theory of Computing*, STOC '93, page 541–550, New York, NY, USA, 1993. Association for Computing Machinery.

[12] Daniel M. Kane and Ryan Williams. Super-linear gate and super-quadratic wire lower bounds for depth-two and depth-three threshold circuits. *CoRR*, abs/1511.07860, 2015.

[13] Moni Naor and Omer Reingold. Number-theoretic constructions of efficient pseudo-random functions. *Journal of the ACM (JACM)*, 51(2):231–262, 2004.

[14] Yuval Peres. Noise stability of weighted majority, 2004.

[15] NP Red'kin. Proof of minimality of circuits consisting of functional elements. *Systems Theory Research: Problemy Kibernetiki*, pages 85–103, 1973.

[16] Kai-Yeung Siu, Vwani P Roychowdhury, and Thomas Kailath. Rational approximation techniques for analysis of neural networks. *IEEE Transactions on Information Theory*, 40(2):455–466, 1994.

[17] Andrew Chi-Chih Yao. Separating the polynomial-time hierarchy by oracles. In *26th Annual Symposium on Foundations of Computer Science (sfcs 1985)*, pages 1–10. IEEE, 1985.