2015 IEEE 56th Annual Symposium on Foundations of Computer Science

An average-case depth hierarchy theorem for Boolean circuits

Benjamin Rossman*, Rocco A. Servedio†, Li-Yang Tan‡
*Simons Institute and NII
†Columbia University
‡Simons Institute

Abstract

We prove an average-case depth hierarchy theorem for Boolean circuits over the standard basis of AND, OR, and NOT gates. Our hierarchy theorem says that for every $d \geq 2$, there is an explicit n-variable Boolean function f, computed by a linear-size depth-d formula, which is such that any depth-$(d-1)$ circuit that agrees with f on $(1/2 + o_n(1))$ fraction of all inputs must have size $\exp(n^{\Omega(1/d)})$. This answers an open question posed by Håstad in his Ph.D. thesis [Hås86b].

Our average-case depth hierarchy theorem implies that the polynomial hierarchy is infinite relative to a random oracle with probability 1, confirming a conjecture of Håstad [Hås86a], Cai [Cai86], and Babai [Bab87]. We also use our result to show that there is no "approximate converse" to the results of Linial, Mansour, Nisan [LMN93] and Boppana [Bop97] on the total influence of constant-depth circuits, thus answering a question posed by Kalai [Kal12] and Hatami [Hat14].

A key ingredient in our proof is a notion of *random projections* which generalize random restrictions.

Keywords

Small-depth circuits; average-case; depth hierarchy theorem; Polynomial Hierarchy; random projections.

I. Introduction

The study of small-depth Boolean circuits is one of the great success stories of complexity theory. The exponential lower bounds against constant-depth AND-OR-NOT circuits [Yao85], [Hås86a], [Raz87], [Smo87] remain among our strongest unconditional lower bounds against concrete models of computation, and the techniques developed to prove these results have led to significant advances in computational learning theory [LMN93], [Man95], pseudorandomness [Nis91], [Baz09], [Raz09], [Bra10], proof complexity [PBI93], [Ajt94], [KPW95], structural complexity [Yao85], [Hås86a], [Cai86], and even algorithm design [Wil14a], [Wil14b], [AWY15].

In addition to *worst-case* lower bounds against small-depth circuits, *average-case* lower bounds, or *correlation bounds*, have also received significant attention. As one recent example, Impagliazzo, Matthews, Paturi [IMP12] and Håstad [Hås14] independently obtained optimal bounds on the correlation of the parity function with small-depth circuits, capping off a long line of work on the problem [Ajt83], [Yao85], [Hås86a], [Cai86], [Bab87], [BIS12]. These results establish strong limits on the computational power of constant-depth circuits, showing that their agreement with the parity function can only be an exponentially small fraction better than that of a constant function.

In this paper we will be concerned with average-case complexity *within* the class of small-depth circuits: our goal is to understand the computational power of depth-d circuits relative to those of strictly smaller depth. Our main result is an *average-case depth hierarchy theorem* for small-depth circuits:

Theorem 1. *Let $2 \leq d \leq \frac{c\sqrt{\log n}}{\log\log n}$, where $c > 0$ is an absolute constant, and Sipser_d be the explicit n-variable read-once monotone depth-d formula described in Section VI. Then any circuit C of depth at most $d-1$ and size at most $S = 2^{n^{\frac{1}{6(d-1)}}}$ over $\{0,1\}^n$ agrees with Sipser_d on at most $(\frac{1}{2} + n^{-\Omega(1/d)}) \cdot 2^n$ inputs.*

(We actually prove two incomparable lower bounds, each of which implies Theorem 1 as a special case. Roughly speaking, the first of these says that Sipser_d cannot be approximated by size-S, depth-d circuits which have significantly smaller bottom fan-in than Sipser_d, and the second of these says that Sipser_d cannot be approximated by size-S, depth-d circuits with a different top-level output gate than Sipser_d.)

Theorem 1 is an average-case extension of the worst-case depth hierarchy theorems of Sipser, Yao, and Håstad [Sip83], [Yao85], [Hås86a], and answers an open problem of Håstad [Hås86a] (which also appears in [Hås86b], [Hås89]). We discuss the background and context for Theorem 1 in Section I-A, and state our two main lower bounds more precisely in Section I-B.

0272-5428/15 $31.00 © 2015 IEEE
DOI 10.1109/FOCS.2015.67

Applications.: We give two applications of our main result, one in structural complexity and the other in the analysis of Boolean functions. First, via a classical connection between small-depth computation and the polynomial hierarchy [FSS81], [Sip83], Theorem 1 implies that the polynomial hierarchy is infinite relative to a random oracle:

Theorem 2. *With probability* 1*, a random oracle* A *satisfies* $\Sigma_d^{\mathrm{P},A} \subsetneq \Sigma_{d+1}^{\mathrm{P},A}$ *for all* $d \in \mathbb{N}$.

This resolves a well-known conjecture in structural complexity, which first appeared in [Hås86a], [Cai86], [Bab87] and has subsequently been discussed in a wide range of surveys [Joh86], [Hem94], [ST95], [HRZ95], [VW97], [Aar], textbooks [DK00], [HO02], and research papers [Hås86b], [Hås89], [Tar89], [For99], [Aar10a]. (Indeed, the results of [Hås86a], [Cai86], [Bab87], along with much of the pioneering work on lower bounds against small-depth circuits in the 1980's, were largely motivated by the aforementioned connection to the polynomial hierarchy.) See Section II for details.

Our second application is a strong negative answer to questions of Kalai and Hatami in the analysis of Boolean functions. Seeking an *approximate converse* to the fundamental results of Linial, Mansour, Nisan [LMN93] and Boppana [Bop97] on the total influence of constant-depth circuits, Kalai asked whether every Boolean function with total influence $\mathrm{polylog}(n)$ can be approximated by a constant-depth circuit of quasipolynomial size [Kal10], [Kal12], [Hat14]. O'Donnell posed a variant of the same question with a more specific quantitative bound on how the size of the approximating circuit depends on its influence and depth [O'D07]. As a consequence of Theorem 1 we obtain the following:

Theorem 3. *There are functions* $d(n) = \omega_n(1)$ *and* $S(n) = \exp((\log n)^{\omega_n(1)})$ *such that there is a monotone* $f : \{0,1\}^n \to \{0,1\}$ *with total influence* $\mathbf{Inf}(f) = O(\log n)$*, but any circuit* C *that has depth* $d(n)$ *and agrees with* f *on at least* $(\frac{1}{2} + o_n(1)) \cdot 2^n$ *inputs in* $\{0,1\}^n$ *must have size greater than* $S(n)$.

Theorem 3 significantly strengthens O'Donnell and Wimmer's counterexample [OW07] to a conjecture of Benjamini, Kalai, and Schramm [BKS99], and shows that the total influence bound for constant-depth circuits of [LMN93], [Bop97] does not admit even a very weak approximate converse. See Section III for details.

A. Previous work

In this subsection we discuss previous work related to our average-case depth hierarchy theorem. We discuss the background and context for our applications, Theorems 2 and 3, in Sections II and III respectively.

Sipser was the first to prove a worst-case depth hierarchy theorem for small-depth circuits [Sip83]. He showed that for every $d \in \mathbb{N}$, there exists a Boolean function $F_d : \{0,1\}^n \to \{0,1\}$ such that F_d is computed by a linear-size depth-d circuit, but any depth-$(d-1)$ circuit computing F_d has size $\Omega(n^{\log^{(3d)} n})$, where $\log^{(i)} n$ denotes the i-th iterated logarithm. The family of functions $\{F_d\}_{d \in \mathbb{N}}$ witnessing this separation are depth-d read-once monotone formulas with alternating layers of AND and OR gates with fan-in $n^{1/d}$ — these came to be known as the *Sipser functions*. Following Sipser's work, Yao claimed an improvement of Sipser's lower bound to $\exp(n^{c_d})$ for some constant $c_d > 0$ [Yao85]. Shortly thereafter Håstad proved a near-optimal separation for (a slight variant of) the Sipser functions:

Theorem 4 (Depth hierarchy of small-depth circuits [Hås86a]; see also [Hås86b], [Hås89])**.** *For every* $d \in \mathbb{N}$*, there exists a Boolean function* $F_d : \{0,1\}^n \to \{0,1\}$ *such that* F_d *is computed by a linear-size depth-*d *circuit, but any depth-*$(d-1)$ *circuit computing* F_d *has size* $\exp(n^{\Omega(1/d)})$.

The parameters of Håstad's theorem were subsequently refined by Cai, Chen, and Håstad [CCH98], and Segerlind, Buss, and Impagliazzo [SBI04]. Prior to the work of Yao and Håstad, Klawe, Paul, Pippenger, and Yannakakis [KPPY84] proved a depth hierarchy theorem for small-depth *monotone* circuits, showing that for every $d \in \mathbb{N}$, depth-$(d-1)$ *monotone* circuits require size $\exp(\Omega(n^{1/(d-1)}))$ to compute the depth-d Sipser function. Klawe et al. also gave an upper bound, showing that every linear-size monotone formula — in particular, the depth-d Sipser function for all $d \in \mathbb{N}$ — can be computed by a depth-k monotone formula of size $\exp(O(k\, n^{1/(k-1)}))$ for all $k \in \mathbb{N}$.

To the best of our knowledge, the first progress towards an *average-case* depth hierarchy theorem for small-depth circuits was made by O'Donnell and Wimmer [OW07]. They constructed a linear-size depth-3 circuit F and proved that any depth-2 circuit that approximates F must have size $2^{\Omega(n/\log n)}$:

Theorem 5 (Theorem 1.9 of [OW07]). *For $w \in \mathbb{N}$ and $n := w2^w$, let $\mathsf{Tribes} : \{0,1\}^n \to \{0,1\}$ be the function computed by a 2^w-term read-once monotone DNF formula where every term has width exactly w. Let $\mathsf{Tribes}^\dagger$ denote its Boolean dual, the function computed by a 2^w-clause read-once monotone CNF formula where every clause has width exactly w, and define the $2n$-variable function $F : \{0,1\}^{2n} \to \{0,1\}$ as*

$$F(x) = \mathsf{Tribes}(x_1, \ldots, x_n) \vee \mathsf{Tribes}^\dagger(x_{n+1}, \ldots, x_{2n}).$$

Then any depth-2 circuit C on $2n$ variables that has size $2^{O(n/\log n)}$ agrees with F on at most a 0.99-fraction of the 2^{2n} inputs. (Note that F is computed by a linear-size depth-3 circuit.)

Our Theorem 1 gives an analogous separation between depth-d and depth-$(d+1)$ for all $d \geq 2$, with $(1/2 - o_n(1))$-inapproximability rather than 0.01-inapproximability. The [OW07] size lower bound of $2^{\Omega(n/\log n)}$ is much larger, in the case $d = 2$, than our $\exp(n^{\Omega(1/d)})$ size bound. However, we recall that achieving a $\exp(\omega(n^{1/(d-1)}))$ lower bound against depth-d circuits for an explicit function, even for worst-case computation, is a well-known and major open problem in complexity theory (see e.g. Chapter §11 of [Juk12] and [Val83], [GW13], [Vio13]). In particular, an extension of the $2^{\Omega(n/\mathrm{polylog}(n))}$-type lower bound of [OW07] to depth 3, even for worst-case computation, would constitute a significant breakthrough.

B. Our main lower bounds

We close this section with precise statements of our two main lower bound results, a discussion of the (near)-optimality of our correlation bounds, and a very high-level overview of our techniques.

Theorem 6 (First main lower bound). *For $2 \leq d \leq \frac{c\sqrt{\log n}}{\log\log n}$, the n-variable Sipser_d function has the following property: Any depth-d circuit $C : \{0,1\}^n \to \{0,1\}$ of size at most $S = 2^{n^{\frac{1}{6(d-1)}}}$ and bottom fan-in $\frac{\log n}{10(d-1)}$ agrees with Sipser_d on at most $(\frac{1}{2} + n^{-\Omega(1/d)}) \cdot 2^n$ inputs.*

Theorem 7 (Second main lower bound). *For $2 \leq d \leq \frac{c\sqrt{\log n}}{\log\log n}$, the n-variable Sipser_d function has the following property: Any depth-d circuit $C : \{0,1\}^n \to \{0,1\}$ of size at most $S = 2^{n^{\frac{1}{6(d-1)}}}$ and the opposite alternation pattern to Sipser_d (i.e. its top-level output gate is OR if Sipser_d's is AND and vice versa) agrees with Sipser_d on at most $(\frac{1}{2} + n^{-\Omega(1/d)}) \cdot 2^n$ inputs.*

Clearly both these results imply Theorem 1 as a special case, since any size-S depth-$(d-1)$ circuit may be viewed as a size-S depth-d circuit satisfying the assumptions of Theorems 6 and 7.

(Near)-optimality of our correlation bounds: For constant d, our main result shows that the depth-d Sipser_d function has correlation at most $(1/2 + n^{-\Omega(1)})$ with any subexponential-size circuit of depth $d-1$. Since Sipser_d is a monotone function, well-known results [BT96] imply that its correlation with some input variable x_i or one of the constant functions 0,1 (trivial approximators of depth at most one) must be at least $(1/2 + \Omega(1/n))$; thus significant improvements on our correlation bound cannot be achieved for this (or for any monotone) function.

What about non-monotone functions? If $\{f_d\}_{d \geq 2}$ is any family of n-variable functions computed by $\mathrm{poly}(n)$-size, depth-d circuits, the "discriminator lemma" of Hajnal et al. [HMP+93] implies that f_d must have correlation at least $(1/2 + n^{-O(1)})$ with one of the depth-$(d-1)$ circuits feeding into its topmost gate. Therefore a "d versus $d-1$" depth hierarchy theorem for correlation $(1/2 + n^{-\omega(1)})$ does not hold.

Our techniques: Our approach is based on *random projections*, a generalization of random restrictions. At a high level, we design a carefully chosen (adaptively chosen) sequence of random projections, and argue that with high probability under this sequence of random projections, (i) any circuit C of the type specified in Theorem 6 or Theorem 7 "collapses," while (ii) the Sipser_d function "retains structure," and (iii) moreover this happens in such a way as to imply that the circuit C must have originally been a very poor approximator for Sipser_d (before the random projections). Each of (i)–(iii) above requires significant work; see Section IV for a much more detailed explanation of our techniques (and of why previous approaches were unable to successfully establish the result).

II. Application #1: Random oracles separate the polynomial hierarchy

A. Background: $\mathsf{PSPACE} \neq \mathsf{PH}$ *relative to a random oracle*

The pioneering work on lower bounds against small-depth circuits in the 1980's was largely motivated by a connection between small-depth computation and the polynomial hierarchy shown by Furst, Saxe, and Sipser [FSS81].

They gave a super-polynomial size lower bound for constant-depth circuits, proving that depth-d circuits computing the n-variable parity function must have size $\Omega(n^{\log^{(3d-6)} n})$, where $\log^{(i)} n$ denotes the i-th iterated logarithm. They also showed that an improvement of this lower bound to super-quasipolynomial for constant-depth circuits (i.e. $\Omega_d\big(2^{(\log n)^k}\big)$ for all constants k) would yield an oracle A such that $\mathsf{PSPACE}^A \neq \mathsf{PH}^A$. Ajtai independently proved a stronger lower bound of $n^{\Omega_d(\log n)}$ [Ajt83]; his motivation came from finite model theory. Yao gave the first super-quasipolynomial lower bounds on the size of constant-depth circuits computing the parity function [Yao85], and shortly after Håstad proved the optimal lower bound of $\exp(\Omega(n^{1/(d-1)}))$ via his influential Switching Lemma [Hås86a].

Yao's relativized separation of PSPACE from PH was improved qualitatively by Cai, who showed that the separation holds even relative to a *random* oracle [Cai86]. Leveraging the connection made by [FSS81], Cai accomplished this by proving *correlation bounds* against constant-depth circuits, showing that constant-depth circuits of sub-exponential size agree with the parity function only on a $(1/2 + o_n(1))$ fraction of inputs. (Independent work of Babai [Bab87] gave a simpler proof of the same relativized separation.)

B. Background: The polynomial hierarchy is infinite relative to some oracle

Together, these results paint a fairly complete picture of the status of the PSPACE versus PH question in relativized worlds: not only does there exist an oracle A such that $\mathsf{PSPACE}^A \neq \mathsf{PH}^A$, this separation holds relative to almost all oracles. A natural next step is to seek analogous results showing that the relativized polynomial hierarchy is infinite; we recall that the polynomial hierarchy being infinite implies $\mathsf{PSPACE} \neq \mathsf{PH}$, and furthermore, this implication relativizes. We begin with the following question, attributed to Albert Meyer in [BGS75]:

Meyer's Question. *Is there a relativized world within which the polynomial hierarchy is infinite? Equivalently, does there exist an oracle A such that $\Sigma_d^{\mathrm{P},A} \subsetneq \Sigma_{d+1}^{\mathrm{P},A}$ for all $d \in \mathbb{N}$?*

Early work on Meyer's question predates [FSS81]. It was first considered by Baker, Gill, and Solovay in their paper introducing the notion of relativization [BGS75], in which they prove the existence of an oracle A such that $\mathsf{P}^A \neq \mathsf{NP}^A \neq \mathsf{coNP}^A$, answering Meyer's question in the affirmative for $d \in \{0, 1\}$. Subsequent work of Baker and Selman proved the $d = 2$ case [BS79]. Following [FSS81], Sipser noted the analogous connection between Meyer's question and circuit lower bounds [Sip83]: to answer Meyer's question in the affirmative, it suffices to exhibit, for every constant $d \in \mathbb{N}$, a Boolean function F_d computable by a depth-d AC^0 circuit such that any depth-$(d-1)$ circuit computing F_d requires super-quasipolynomial size. (This is a significantly more delicate task than proving super-quasipolynomial size lower bounds for the parity function; see Section IV for a detailed discussion.) Sipser also constructed a family of Boolean functions for which he proved an n versus $\Omega(n^{\log^{(3d)} n})$ separation — these came to be known as the *Sipser functions*, and they play the same central role in Meyer's question as the parity function does in the relativized PSPACE versus PH problem.

As discussed in the introduction (see Theorem 4), Håstad gave the first proof of a near-optimal n versus $\exp(n^{\Omega(1/d)})$ separation for the Sipser functions [Hås86a], obtaining a strong depth hierarchy theorem for small-depth circuits and answering Meyer's question in the affirmative for all $d \in \mathbb{N}$.

C. This work: The polynomial hierarchy is infinite relative to a random oracle

Given Håstad's result, a natural goal is to complete our understanding of Meyer's question by showing that the polynomial hierarchy is not just infinite with respect to *some* oracle, but in fact with respect to *almost all* oracles. Indeed, in [Hås86a], [Hås86b], [Hås89], Håstad poses the problem of extending his result to show this as an open question:

Question 1 (Meyer's Question for Random Oracles [Hås86a], [Hås86b], [Hås89])**.** *Is the polynomial hierarchy infinite relative to a random oracle? Equivalently, does a random oracle A satisfy $\Sigma_d^{\mathrm{P},A} \subsetneq \Sigma_{d+1}^{\mathrm{P},A}$ for all $d \in \mathbb{N}$?*

Question 1 also appears as the main open problem in [Cai86], [Bab87]; as mentioned above, an affirmative answer to Question 1 would imply Cai and Babai's result showing that $\mathsf{PSPACE}^A \neq \mathsf{PH}^A$ relative to a random oracle A. Further motivation for studying Question 1 comes from a surprising result of Book, who proved that the *unrelativized* polynomial hierarchy collapses if it collapses relative to a random oracle [Boo94]. Over the years Question 1 has been discussed in a wide range of surveys [Joh86], [Hem94], [ST95], [HRZ95], [VW97], [Aar], textbooks [DK00], [HO02], and research papers [Hås86b], [Hås89], [Tar89], [For99], [Aar10a].

Our work: As a corollary of our main result (Theorem 1) — an *average-case* depth hierarchy theorem for small-depth circuits — we answer Question 1 in the affirmative for all $d \in \mathbb{N}$:

Theorem 2. *The polynomial hierarchy is infinite relative to a random oracle: with probability* 1*, a random oracle* A *satisfies* $\Sigma_d^{\mathrm{P},A} \subsetneq \Sigma_{d+1}^{\mathrm{P},A}$ *for all* $d \in \mathbb{N}$.

Prior to our work, the $d \in \{0,1\}$ cases were proved by Bennett and Gill in their paper initiating the study of random oracles [BG81]. Motivated by the problem of obtaining relativized separations in quantum structural complexity, Aaronson recently showed that a random oracle A separates Π_2^{P} from P^{NP} [Aar10b], [Aar10a]; he conjectures in [Aar10a] that his techniques can be extended to resolve the $d = 2$ case of Theorem 2. We observe that O'Donnell and Wimmer's techniques (Theorem 5 in our introduction) can be used to prove the $d = 2$ case [OW07], though the authors of [OW07] do not discuss this connection to the relativized polynomial hierarchy in their paper.

	$\mathsf{PSPACE}^A \neq \mathsf{PH}^A$	$\Sigma_d^{\mathrm{P},A} \subsetneq \Sigma_{d+1}^{\mathrm{P},A}$ for all $d \in \mathbb{N}$
Connection to lower bounds for constant-depth circuits	[FSS81]	[Sip83]
Hard function(s)	Parity	Sipser functions
Relative to *some* oracle A	[Yao85], [Hås86a]	[Yao85], [Hås86a]
Relative to *random* oracle A	[Cai86], [Bab87]	**This work**

Table I: Previous work and our result on the relativized polynomial hierarchy

We refer the reader to Chapter §7 of Håstad's thesis [Hås86b] for a detailed exposition (and complete proofs) of the aforementioned connections between small-depth circuits and the polynomial hierarchy (in particular, for the proof of how Theorem 2 follows from Theorem 1).

III. Application #2: No approximate converse to Boppana–Linial–Mansour–Nisan

The famous result of Linial, Mansour, and Nisan gives strong bounds on Fourier concentration of small-depth circuits [LMN93]. As a corollary, they derive an upper bound on the total influence of small-depth circuits, showing that depth-d size-S circuits have total influence $(O(\log S))^d$. (We remind the reader that the total influence of an n-variable Boolean function f is $\mathbf{Inf}(f) := \sum_{i=1}^n \mathbf{Inf}_i(f)$, where $\mathbf{Inf}_i(f)$ is the probability that flipping coordinate $i \in [n]$ of a uniform random input from $\{0,1\}^n$ causes the value of f to change.) This was subsequently sharpened by Boppana via a simpler and more direct proof [Bop97]:

Theorem 8 (Boppana, Linial–Mansour–Nisan)**.** *Let* $f : \{0,1\}^n \to \{0,1\}$ *be a computed by a size-*S *depth-*d *circuit. Then* $\mathbf{Inf}(f) = (O(\log S))^{d-1}$.

(We note that Boppana's bound is asymptotically tight by considering the parity function.) Several researchers have asked whether an *approximate converse* of some sort holds for Theorem 8:

> *If* $f : \{0,1\}^n \to \{0,1\}$ *has low total influence, is it the case that* f *can be approximated to high accuracy by a small constant-depth circuit?*

A result of this flavor, taken together with Theorem 8, would yield an elegant characterization of Boolean functions with low total influence. In this section we formulate a very weak approximate converse to Theorem 8 and show, as a consequence of our main result (Theorem 1), that even this weak converse does not hold.

A. Background: BKS conjecture and O'Donnell–Wimmer's counterexample

An approximate converse to Theorem 8 was first conjectured by Benjamini, Kalai, and Schramm, with a very specific quantitative bound on how the size of the approximating circuit depends on its influence and depth [BKS99] (the conjecture also appears in the surveys [Kal00], [KS05]). They posed the following:

Benjamini–Kalai–Schramm (BKS) Conjecture. *For every* $\varepsilon > 0$ *there is a constant* $K = K(\varepsilon)$ *such that the following holds: Every monotone* $f : \{0,1\}^n \to \{0,1\}$ *can be* ε*-approximated by a depth-*d *circuit of size at most*

$$\exp\left((K \cdot \mathbf{Inf}(f))^{1/(d-1)}\right)$$

for some $d \geq 2$.

(We associate a circuit with the Boolean function that it computes, and we say that a circuit ε-*approximates* a Boolean function f if it agrees with f on all but an ε-fraction of all inputs.) If true, the BKS conjecture would give a quantitatively strong converse to Theorem 8 for monotone functions.[1] In addition, it would have important implications for the study of threshold phenomena in Erdös–Rényi random graphs, which is the context in which Benjamini, Kalai, and Schramm made their conjecture; we refer the reader to [BKS99] and Section 1.4 of [OW07] for a detailed discussion of this connection. However, the BKS conjecture was disproved by O'Donnell and Wimmer [OW07]. Their result (Theorem 5 in our introduction) disproves the case $d = 2$ of the BKS conjecture, and the case $d > 2$ is disproved by an easy argument which [OW07] give.

B. This work: Disproving a weak variant of the BKS conjecture for constant depth

A significantly weaker variant of the BKS conjecture for constant depth is the following:

Conjecture 1. *For every $\varepsilon > 0$ there is a $d = d(\varepsilon)$ and $K_1 = K_1(\varepsilon), K_2 = K_2(\varepsilon)$ such that the following holds: Every monotone $f : \{0,1\}^n \to \{0,1\}$ can be ε-approximated by a depth-d circuit of size at most*

$$\exp\left((K_1 \cdot \mathbf{Inf}(f))^{K_2}\right).$$

Conjecture 1 is incomparable to the BKS conjecture; on the one hand, it stipulates that d depends only on ε and not on n, but on the other hand it allows K_2 to depends on ε (and hence on d) in an arbitrary way. The [OW07] counterexample to the BKS conjecture does not disprove Conjecture 1; indeed, the function f that [OW07] construct and analyze is computed by a depth-3 circuit of size $O(n)$.[2] Observe that Conjecture 1, if true, would yield the following rather appealing consequence: every monotone $f : \{0,1\}^n \to \{0,1\}$ with total influence at most $\mathrm{polylog}(n)$ can be approximated to any constant accuracy by a quasipolynomial-size, constant-depth circuit (where both the constant in the quasipolynomial size bound and the constant depth of the circuit may depend on the desired accuracy).

Following O'Donnell and Wimmer's disproof of the BKS conjecture, several researchers have posed questions similar in spirit to Conjecture 1. O'Donnell asked if the BKS conjecture is true if the bound on the size of the approximating circuit is allowed to be $\exp\left((K \cdot \mathbf{Inf}(f))^{1/(d-2)}\right)$ instead of $\exp\left((K \cdot \mathbf{Inf}(f))^{1/(d-1)}\right)$ [O'D07]. This is a weaker statement than the original BKS conjecture (in particular, it is not ruled out by the counterexample of [OW07]), but still significantly stronger than Conjecture 1 for constant values of d. Subsequently Kalai asked if Boolean functions with total influence $\mathrm{polylog}(n)$ (resp. $O(\log n)$) can be approximated by constant-depth circuits of quasipolynomial size (resp. AC^0) [Kal12] (see also [Kal10] where he states a qualitative version). Kalai's question is a variant of Conjecture 1 in which f is allowed to be non-monotone, but $\mathbf{Inf}(f)$ is only allowed to be $\mathrm{polylog}(n)$; furthermore, $K_2(\varepsilon)$ is only allowed to be 1 if $\mathbf{Inf}(f) = O(\log n)$. Finally, H. Hatami recently restated the $\mathbf{Inf}(f) = O(\log n)$ case of Kalai's question:

Problem 4.6.3 of [Hat14]. *Is it the case that for every $\varepsilon, C > 0$, there are constants d, k such that for every $f : \{0,1\}^n \to \{0,1\}$ with $\mathbf{Inf}(f) \leq C \log n$, there is a size-$n^k$, depth-$d$ circuit which ε-approximates f?*

Our work: As a corollary of our main result (Theorem 1), we show that Conjecture 1 is false even for (suitable choices of) $\varepsilon = \frac{1}{2} - o_n(1)$. Our counterexample also provides a strong negative answer to O'Donnell's and Kalai–Hatami's versions of Conjecture 1. We prove the following:

Theorem 3. *Conjecture 1 is false. More precisely, there is a monotone $f : \{0,1\}^n \to \{0,1\}$ and a $\delta(n) = o_n(1)$ such that $\mathbf{Inf}(f) = O(\log n)$ but any circuit of depth $d(n) = \sqrt{\log \log n}$ that agrees with f on $(\frac{1}{2} + \delta(n))$ fraction of all inputs must have size at least $S(n) = 2^{2^{\tilde{\Omega}\left(2^{\sqrt{\log \log n}}\right)}}$.*

Proof of Theorem 3 assuming Theorem 1: Consider the monotone Boolean function $f : \{0,1\}^n \to \{0,1\}$ corresponding to Sipser_d of Theorem 1 defined over the first $m = 2^{2^{\lfloor \sqrt{\log \log n} \rfloor}}$ variables, and of depth $d =$

[1] We remark that although the BKS conjecture was stated for monotone Boolean functions, it seems that (a priori) it could have been true for all Boolean functions: prior to [OW07], we are not aware of any counterexample to the BKS conjecture even if f is allowed to be non-monotone.

[2] As with the BKS conjecture, prior to our work we are not aware of any counterexample to Conjecture 1 even if f is allowed to be non-monotone.

$\lfloor \log \log m \rfloor + 1 = \lfloor \sqrt{\log \log n} \rfloor + 1$. By Boppana's theorem (Theorem 8), we have that

$$\mathbf{Inf}(f) = O(\log m)^{d-1} = O\left(2^{\lfloor \sqrt{\log \log n} \rfloor}\right)^{\lfloor \sqrt{\log \log n} \rfloor} = O(\log n).$$

On the other hand, our main theorem (Theorem 1) implies that even circuits of depth $d-1 = \lfloor \sqrt{\log \log n} \rfloor$ which agree with f on $(\frac{1}{2} + \delta(n))$ fraction of all inputs, where $\delta(n) = 2^{-\Omega(2^{\lfloor \sqrt{\log \log n} \rfloor} / \lfloor \sqrt{\log \log n} \rfloor)}$, must have size at least

$$S(n) = 2^{m^{\Omega(1/d)}} = 2^{\left(2^{2\sqrt{\log \log n}}\right)^{\Omega(1/\sqrt{\log \log n})}} = 2^{2^{\tilde{\Omega}\left(2\sqrt{\log \log n}\right)}}.$$

■

IV. Our techniques

The method of random restrictions dates back to Subbotovskaya [Sub61] and continues to be an indispensable technique in circuit complexity. Focusing only on small-depth circuits, we mention that the random restriction method is the common essential ingredient underlying the landmark lower bounds discussed in the previous sections [FSS81], [Ajt83], [Sip83], [Yao85], [Hås86a], [Cai86], [Bab87], [IMP12], [Hås14].

We begin in Section IV-A by describing the general framework for proving worst- and average-case lower bounds against small-depth circuits via the random restriction method. Within this framework, we sketch the now-standard proof of correlation bounds for the parity function based on Håstad's Switching Lemma. We also recall why the lemma is not well-suited for proving a depth hierarchy theorem for small-depth circuits, hence necessitating the "blockwise variant" of the lemma that Håstad developed and applied to prove his (worst-case) depth hierarchy theorem. In Section IV-B we highlight the difficulties that arise in extending Håstad's depth hierarchy theorem to the average-case, and how our techniques — specifically, the notion of random *projections* — allow us to overcome these difficulties.

A. Background: Lower bounds via random restrictions

Suppose we would like to show that a *target function* $f : \{0,1\}^n \to \{0,1\}$ has small correlation with any size-S depth-d *approximating circuit* C under the uniform distribution $\mathcal{U}$ over $\{0,1\}^n$. A standard approach is to construct a series of random restrictions $\{\mathcal{R}_k\}_{k \in \{2,\ldots,d\}}$ satisfying three properties:

- **Property 1: Approximator C simplifies.** The randomly-restricted circuit $C \restriction \boldsymbol{\rho}^{(d)} \cdots \boldsymbol{\rho}^{(2)}$, where $\boldsymbol{\rho}^{(k)} \leftarrow \mathcal{R}_k$ for $2 \le k \le d$, should "collapse to a simple function" with high probability. This is typically shown via iterative applications of an appropriate "Switching Lemma for the $\mathcal{R}_k$'s", which shows that each random restriction $\boldsymbol{\rho}^{(k)}$ decreases the depth of the circuit $C \restriction \boldsymbol{\rho}^{(d)} \cdots \boldsymbol{\rho}^{(k-1)}$ by one with high probability. The upshot is that while C is a depth-d size-S circuit, $C \restriction \boldsymbol{\rho}^{(d)} \cdots \boldsymbol{\rho}^{(2)}$ will be a small-depth decision tree, a "simple function", with high probability.
- **Property 2: Target f retains structure.** In contrast with the approximating circuit, the target function f should (roughly speaking) be resilient against the random restrictions $\boldsymbol{\rho}^{(k)} \leftarrow \mathcal{R}_k$. While the precise meaning of "resilient" depends on the specific application, the key property we need is that $f \restriction \boldsymbol{\rho}^{(d)} \cdots \boldsymbol{\rho}^{(2)}$ will with high probability be a "well-structured" function that is uncorrelated with any small-depth decision tree.

Together, these two properties imply that random restrictions of f and C are uncorrelated with high probability. Note that this already yields *worst-case* lower bounds, showing that $f : \{0,1\}^n \to \{0,1\}$ cannot be computed exactly by C. To obtain correlation bounds, we need to translate such a statement into the fact that f and C *themselves* are uncorrelated. For this we need the third key property of the random restrictions:

- **Property 3: Composition of $\mathcal{R}_k$'s completes to $\mathcal{U}$.** Evaluating a Boolean function $h : \{0,1\}^n \to \{0,1\}$ on a random input $\mathbf{X} \leftarrow \mathcal{U}$ is equivalent to first applying random restrictions $\boldsymbol{\rho}^{(d)}, \ldots, \boldsymbol{\rho}^{(2)}$ to h, and then evaluating the randomly-restricted function $h \restriction \boldsymbol{\rho}^{(d)} \cdots \boldsymbol{\rho}^{(2)}$ on $\mathbf{X}' \leftarrow \mathcal{U}$.

Correlation bounds for parity: For uniform-distribution correlation bounds against constant-depth circuits computing the parity function, the random restrictions are all drawn from $\mathcal{R}(p)$, the "standard" random restriction which independently sets each free variable to 0 with probability $\frac{1}{2}(1-p)$, to 1 with probability $\frac{1}{2}(1-p)$, and keeps it free with probability p. The main technical challenge arises in proving that Property 1 holds — this is

precisely Håstad's Switching Lemma — whereas Properties 2 and 3 are straightforward to show. For the second property, we note that

$$\mathsf{Parity}_n \restriction \rho \equiv \pm\,\mathsf{Parity}(\rho^{-1}(*)) \quad \text{for all restrictions } \rho \in \{0,1,*\}^n,$$

and so $\mathsf{Parity}_n \restriction \boldsymbol{\rho}^{(d)} \cdots \boldsymbol{\rho}^{(2)}$ computes the parity of a random subset $\mathbf{S} \subseteq [n]$ of coordinates (or its negation). With an appropriate choice of the $*$-probability p we have that $|\mathbf{S}|$ is large with high probability; recall that $\pm\,\mathsf{Parity}_k$ (the k-variable parity function or its negation) has zero correlation with any decision tree of depth at most $k-1$. For the third property, we note that for all values of $p \in (0,1)$, a random restriction $\boldsymbol{\rho} \leftarrow \mathcal{R}(p)$ specifies a uniform random subcube of $\{0,1\}^n$ (of dimension $|\boldsymbol{\rho}^{-1}(*)|$). Therefore, the third property is a consequence of the simple fact that a uniform random point within a uniform random subcube is itself a uniform random point from $\{0,1\}^n$.

Håstad's blockwise random restrictions: With the above framework in mind, we notice a conceptual challenge in proving AC^0 depth hierarchy theorems via the random restriction method: even focusing only on the worst-case (i.e. ignoring Property 3), the random restrictions $\mathcal{R}_k$ will have to satisfy Properties 1 and 2 with the target function f being *computable in* AC^0. This is a significantly more delicate task than (say) proving $\mathsf{Parity} \notin \mathsf{AC}^0$ since, roughly speaking, in the latter case the target function $f \equiv \mathsf{Parity}$ is "much more complex" than the circuit $C \in \mathsf{AC}^0$ to begin with. In an AC^0 depth hierarchy theorem, *both* the target f and the approximating circuit C are constant-depth circuits; the target f is "more complex" than C in the sense that it has larger circuit depth, but this is offset by the fact that the circuit size of C is allowed to be exponentially larger than that of f (as is the case in both Håstad's and our theorem). We refer the reader to Chapter §6.2 of Hastad's thesis [Hås86b] which contains a discussion of this very issue.

Håstad overcomes this difficulty by replacing the "standard" random restrictions $\mathcal{R}(p)$ with random restrictions *specifically suited to Sipser functions being the target*: his "blockwise" random restrictions are designed so that (1) they reduce the depth of the formula computing the Sipser function by one, but otherwise essentially preserve the rest of its structure, and yet (2) a switching lemma still holds for any circuit with sufficiently small bottom fan-in. These correspond to Properties 2 and 1 respectively. However, unlike $\mathcal{R}(p)$, Håstad's blockwise random restrictions are not independent across coordinates and do not satisfy Property 3: their composition does not complete to the uniform distribution $\mathcal{U}$ (and indeed it does not complete to any product distribution). This is why Håstad's construction establishes a worst-case rather than average-case depth hierarchy theorem.

B. Our main technique: Random projections

The crux of the difficulty in proving an average-case AC^0 depth hierarchy theorem therefore lies in designing random restrictions that satisfy Properties 1, 2, and 3 simultaneously, for a target f in AC^0 and an arbitrary approximating circuit C of smaller depth but possibly exponentially larger size. To recall, the "standard" random restrictions $\mathcal{R}(p)$ satisfy Properties 1 and 3 but not 2, and Håstad's blockwise variant satisfies Properties 1 and 2 but not 3.

In this paper we overcome this difficulty with *projections*, a generalization of restrictions. Given a set of formal variables $\mathcal{X} = \{x_1, \ldots, x_n\}$, a restriction ρ either fixes a variable x_i (i.e. $\rho(x_i) \in \{0,1\}$) or keeps it alive (i.e. $\rho(x_i) = x_i$, often denoted by $*$). A *projection*, on the other hand, either fixes x_i or maps it to a variable y_j from a possibly different space of formal variables $\mathcal{Y} = \{y_1, \ldots, y_{n'}\}$. Restrictions are therefore a special case of projections where $\mathcal{Y} \equiv \mathcal{X}$, and each x_i can only be fixed or mapped to itself. (See Definition 4 for precise definitions.) Our arguments crucially employ projections in which $\mathcal{Y}$ is smaller than $\mathcal{X}$, and where moreover each x_i is only mapped to a specific element y_j where j depends on i in a carefully designed way that depends on the structure of the formula computing the Sipser function. Such "collisions", where blocks of distinct formal variables in $\mathcal{X}$ are mapped to the same new formal variable $y_i \in \mathcal{Y}$, play a crucial role in our approach. (We remark that ours is not the first work to consider such a generalization of restrictions. Random projections are also used in the work of Impagliazzo and Segerlind, which establishes lower bounds against constant-depth Frege systems with counting axioms in proof complexity [IS01].)

At a high level, our overall approach is structured around a sequence $\boldsymbol{\Psi}$ of (*adaptively chosen*) random projections satisfying Properties 1, 2, and 3 simultaneously, with the target f being Sipser, a slight variant of the Sipser function which we define in Section VI. We briefly outline how we establish each of the three properties (it will be more natural for us to prove them in a slightly different order from the way they are listed in Section IV-A):

– **Property 3: $\mathbf{\Psi}$ completes to the uniform distribution.** Like Håstad's blockwise random restrictions (and unlike the "standard" random restrictions $\mathcal{R}(p)$), the distributions of our random projections are not independent across coordinates: they are carefully correlated in a way that depends on the structure of the formula computing Sipser. As discussed above, there is an inherent tension between the need for such correlations on one hand (to ensure that Sipser "retains structure"), and the requirement that their composition completes to the uniform distribution on the other hand (to yield uniform-distribution correlation bounds). We overcome this difficulty with our notion of projections: in Section VIII of the full version we prove that the composition $\mathbf{\Psi}$ of our sequence of random projections completes to the uniform distribution (despite the fact that every one of the individual random projections comprising $\mathbf{\Psi}$ is highly-correlated among coordinates.)
– **Property 1: Approximator C simplifies.** Next we prove that approximating circuits C of the types specified in our main lower bounds (Theorems 6 and 7) "collapse to a simple function" with high probability under our sequence $\mathbf{\Psi}$ of random projections. Following the standard "bottom-up" approach to proving lower bounds against small-depth circuits, we establish this by arguing that each of the individual random projections comprising $\mathbf{\Psi}$ "contributes to the simplification" of C by reducing its depth by (at least) one.
  More precisely, we prove a *projection switching lemma*, showing that a small-width DNF or CNF "switches" to a small-depth decision tree with high probability under our random projections. (The depth reduction of C follows by applying this lemma to every one of its bottom-level depth-2 subcircuits.) Recall that the random projection of a depth-2 circuit over a set of formal variables $\mathcal{X}$ yields a function over a new set of formal variables $\mathcal{Y}$, and in our case $\mathcal{Y}$ is significantly smaller than $\mathcal{X}$. In addition to the structural simplification that results from setting variables to constants (as in Håstad's Switching Lemma for random *restrictions*), the proof of our projection switching lemma also crucially exploits the additional structural simplification that results from distinct variables in $\mathcal{X}$ being mapped to the same variable in $\mathcal{Y}$.
– **Property 2: Target** Sipser **retains structure.** Like Håstad's blockwise random restrictions, our random projections are defined with the target function Sipser in mind; in particular, they are carefully designed so as to ensure that Sipser "retains structure" with high probability under their composition $\mathbf{\Psi}$.
  We define the notion of a "typical" outcome of our random projections, and prove that with high probability *all* the individual projections comprising $\mathbf{\Psi}$ are typical. (Since our sequence of random projections is chosen adaptively, this requires a careful definition of typicality to facilitate an inductive argument showing that our definition "bootstraps" itself.) Next, we show that typical projections have a "very limited and well-controlled" effect on the structure of Sipser; equivalently, Sipser is resilient against typical projections. Together, these show that with high probability, Sipser reduces under $\mathbf{\Psi}$ to a "well-structured" formula, in sharp contrast with our results above showing that the approximator "collapses to a simple function" with high probability under $\mathbf{\Psi}$.

We remark that the notion of random projections plays a key role in ensuring all three properties above. (We give a more detailed overview of our proof in Section VII-C after setting up the necessary terminology and definitions in the next two sections.)

V. Preliminaries

A. Notation

A DNF is an OR of ANDs (terms) and a CNF is an AND of ORs (clauses). The *width* of a DNF (respectively, CNF) is the maximum number of variables that occur in any one of its terms (respectively, clauses). We will assume throughout that our circuits are *alternating*, meaning that every root-to-leaf path alternates between AND gates and OR gates, and *layered*, meaning that for every gate G, every root-to-G path has the same length. By a standard conversion, every depth-d circuit is equivalent to a depth-d alternating layered circuit with only a modest increase in size (which is negligible given the slack on our analysis). The size of a circuit is its number of gates, and the depth of a circuit is the length of its longest root-to-leaf path.

For $p \in [0,1]$ and symbols $\bullet, \circ$, we write "$\{\bullet_p, \circ_{1-p}\}$" to denote the distribution over $\{\bullet, \circ\}$ which outputs $\bullet$ with probability p and $\circ$ with probability $1-p$. We write "$\{\bullet_p, \circ_{1-p}\}^k$" to denote the product distribution over $\{\bullet, \circ\}^k$ in which each coordinate is distributed independently according to $\{\bullet_p, \circ_{1-p}\}$. We write "$\{\bullet_p, \circ_{1-p}\}^k \setminus \{\bullet\}^k$" to denote the product distribution conditioned on not outputting $\{\bullet\}^k$.

Given $\tau \in \{0,1,*\}^{A \times [\ell]}$ and $a \in A$, we write τ_a to denote the ℓ-character string $(\tau_{a,i})_{i \in [\ell]} \in \{0,1,*\}^{[\ell]}$, and we sometimes refer to this as the "a-th block of τ."

Throughout the paper we use boldfaced characters such as $\boldsymbol{\rho}$, $\mathbf{X}$, etc. to denote random variables. We write "$a = b \pm c$" as shorthand to denote that $a \in [b-c, b+c]$, and similarly $a \neq b \pm c$ to denote that $a \notin [b-c, b+c]$. For a positive integer k we write "$[k]$" to denote the set $\{1, \ldots, k\}$.

The *bias* of a Boolean function f under an input distribution $\mathbf{Z}$ is defined as

$$\mathrm{bias}(f, \mathbf{Z}) := \min\left\{ \mathop{\mathbf{Pr}}_{\mathbf{Z}}[f(\mathbf{Z}) = 0], \mathop{\mathbf{Pr}}_{\mathbf{Z}}[f(\mathbf{Z}) = 1] \right\}.$$

B. Restrictions and random restrictions

Definition 1 (Restriction)**.** *A* restriction *ρ of a finite base set $\{x_\alpha\}_{\alpha \in \Omega}$ of Boolean variables is a string $\rho \in \{0,1,*\}^\Omega$. (We sometimes equivalently view a restriction ρ as a function $\rho : \Omega \to \{0,1,*\}$.) Given a function $f : \{0,1\}^\Omega \to \{0,1\}$ and restriction $\rho \in \{0,1,*\}^\Omega$, the* ρ-restriction *of f is the function $(f \restriction \rho) : \{0,1\}^\Omega \to \{0,1\}$ where*

$$(f \restriction \rho)(x) = f(x \restriction \rho), \quad \text{and } (x \restriction \rho)_\alpha := \begin{cases} x_\alpha & \text{if } \rho_\alpha = * \\ \rho_\alpha & \text{otherwise} \end{cases} \quad \text{for all } \alpha \in \Omega.$$

Given a distribution $\mathcal{R}$ over restrictions $\{0,1,*\}^\Omega$ the $\mathcal{R}$-random restriction *of f is the random function $f \restriction \boldsymbol{\rho}$ where $\boldsymbol{\rho} \leftarrow \mathcal{R}$.*

Definition 2 (Refinement)**.** *Let $\rho, \tau \in \{0,1,*\}^\Omega$ be two restrictions. We say that τ is a* refinement *of ρ if $\rho^{-1}(1) \subseteq \tau^{-1}(1)$ and $\rho^{-1}(0) \subseteq \tau^{-1}(0)$, i.e. every variable x_α that is set to 0 or 1 by ρ is set in the same way by τ (and τ may set additional variables to 0 or 1 that ρ does not set).*

Definition 3 (Composition)**.** *Let $\rho, \rho' \in \{0,1,*\}^\Omega$ be two restrictions. Their* composition*, denoted $\rho\rho' \in \{0,1,*\}^\Omega$, is the restriction defined by*

$$(\rho\rho')_\alpha = \begin{cases} \rho_\alpha & \text{if } \rho_\alpha \in \{0,1\} \\ \rho'_\alpha & \text{otherwise.} \end{cases}$$

Note that $\rho\rho'$ is a refinement of ρ.

C. Projections and random projections

A key ingredient in this work is the notion of *random projections* which generalize random restrictions. Throughout the paper we will be working with functions over spaces of formal variables that are partitioned into disjoint blocks of some length ℓ (see Section VI for a precise description of these spaces). In other words, our functions will be over spaces of formal variables that can be described as $\mathcal{X} = \{x_{a,i} \colon a \in A, i \in [\ell]\}$, where we refer to $x_{a,i}$ as the i-th variable in the a-th block. We associate with each such space $\mathcal{X}$ a smaller space $\mathcal{Y} = \{y_a \colon a \in A\}$ containing a new formal variable for each block of $\mathcal{X}$. Given a function f over $\mathcal{X}$, the *projection* of f yields a function over $\mathcal{Y}$, and the *random projection* of f is the projection of a random restriction of f (which again is a function over $\mathcal{Y}$). Formally, we have the following definition:

Definition 4 (Projection)**.** *The* projection operator proj *acts on functions $f : \{0,1\}^{A \times [\ell]} \to \{0,1\}$ as follows. The* projection of *f is the function $(\mathrm{proj}\, f) : \{0,1\}^A \to \{0,1\}$ defined by*

$$(\mathrm{proj}\, f)(y) = f(x) \quad \text{where } x_{a,i} = y_a \text{ for all } a \in A \text{ and } i \in [\ell].$$

Given a restriction $\rho \in \{0,1,*\}^{A \times [\ell]}$, the ρ-projection of *f is the function $(\mathrm{proj}_\rho\, f) : \{0,1\}^A \to \{0,1\}$ defined by*

$$(\mathrm{proj}_\rho\, f)(y) = f(x) \quad \text{where } x_{a,i} = \begin{cases} y_a & \text{if } \rho_{a,i} = * \\ \rho_{a,i} & \text{otherwise} \end{cases} \quad \text{for all } a \in A \text{ and } i \in [\ell].$$

Equivalently, $(\mathrm{proj}_\rho\, f) \equiv (\mathrm{proj}\,(f \restriction \rho))$. Given a distribution $\mathcal{R}$ over restrictions in $\{0,1,*\}^{A \times [\ell]}$, the associated random projection operator is $\mathrm{proj}_{\boldsymbol{\rho}}$ where $\boldsymbol{\rho} \leftarrow \mathcal{R}$, and for $f : \{0,1\}^{A \times [\ell]} \to \{0,1\}$ we call $\mathrm{proj}_{\boldsymbol{\rho}}\, f$ its $\mathcal{R}$-random projection*.*

Note that when $\ell = 1$, the spaces $\mathcal{X}$ and $\mathcal{Y}$ are identical and our definitions of a ρ-projection and $\mathcal{R}$-random projection coincide exactly with that of a ρ-restriction and $\mathcal{R}$-random restriction in Definition 1 (in this case the projection operator proj is simply the identity operator).

Remark 9. The following interpretation of the projection operator will be useful for us. Let f be a function over $\mathcal{X}$, and consider its representation as a circuit C (or decision tree) accessing the formal variables $x_{a,i}$ in $\mathcal{X}$. The projection of f is the function computed by the circuit C', where C' is obtained from C by replacing every occurrence of $x_{a,i}$ in C by y_a for all $a \in A$ and $i \in [\ell]$. Note that this may result in a significant simplification of the circuit: for example, an AND gate (OR gate, respectively) in C that access both $x_{a,i}$ and $\overline{x}_{a,j}$ for some $a \in A$ and $i, j \in [\ell]$ will access both y_a and $\overline{y}_a$ in C', and therefore can be simplified and replaced by the constant 0 (1, respectively). This is a fact we will exploit in the proof of our projection switching lemma.

VI. The Sipser function and its basic properties

For $2 \leq d \in \mathbb{N}$, in this subsection we define the depth-d monotone n-variable read-once Boolean formula Sipser_d and establish some of its basic properties. The Sipser_d function is very similar to the depth-d formula considered by Håstad [Hås86b]; the only difference is that the fan-ins of the gates in the top and bottom layers have been slightly adjusted, essentially so as to ensure that the formula is very close to balanced between the two output values 0 and 1 (note that such balancedness is a prerequisite for any $(1/2 - o_n(1))$-inapproximability result.) The Sipser_d formula is defined in terms of an integer parameter m; in all our results this is an asymptotic parameter that approaches $+\infty$, so m should be thought of as "sufficiently large" throughout the paper.

Every leaf of Sipser_d occurs at the same depth (distance from the root) d; there are exactly n leaves (n will be defined below) and each variable occurs at precisely one leaf. The formula is *alternating*, meaning that every root-to-leaf path alternates between AND gates and OR gates; all of the gates that are adjacent to input variables (i.e. the depth-$(d-1)$ gates) are AND gates, so the root is an OR gate if d is even and is an AND gate if d is odd. The formula is also *depth-regular*, meaning that for each depth (distance from the root) $0 \leq k \leq d-1$, all of the depth-k gates have the same fan-in. Hence to completely specify the Sipser_d formula it remains only to specify the fan-in sequence $w_0, \ldots, w_{d-1}$, where w_k is the fan-in of every gate at depth k. These fan-ins are as follows:

– The bottommost fan-in is

$$w_{d-1} := m. \tag{1}$$

We define

$$p := 2^{-w_{d-1}} = 2^{-m}, \tag{2}$$

and we observe that p is the probability that a depth-$(d-1)$ AND gate is satisfied by a uniform random choice of $\mathbf{X} \leftarrow \{0_{1/2}, 1_{1/2}\}^n$.

– For each value $1 \leq k \leq d-2$, the value of w_k is $w_k = w$ where

$$w := \lfloor m2^m / \log(e) \rfloor. \tag{3}$$

– The value w_0 is defined to be

$$w_0 := \text{the smallest integer such that } (1-t_1)^{qw_0} \text{ is at most } \frac{1}{2}, \tag{4}$$

where t_1 and q will be defined in Section VII-A, see specifically Equations (8) and (7). Roughly speaking, w_0 is chosen so that the overall formula is essentially balanced under the uniform distribution (i.e. Sipser_d satisfies (6) below); see (9) and the discussion thereafter.

The number of input variables n for Sipser_d is $n = \prod_{k=0}^{d-1} w_k = w^{d-2} w_{d-1} w_0$. The estimates for t_1 and q given in (10) imply that $w_0 = 2^m \ln(2) \cdot (1 \pm o_m(1))$, so we have that

$$n = \frac{1 \pm o_m(1)}{\log e} \cdot \left(\frac{m2^m}{\log e}\right)^{d-1}. \tag{5}$$

We note that for the range of values $2 \leq d \leq \frac{c\sqrt{\log n}}{\log\log n}$ that we consider in this paper, a direct (but somewhat tedious) analysis implies that the Sipser_d function is indeed essentially balanced, or more precisely, that it satisfies

$$\mathop{\mathbf{Pr}}_{\mathbf{X} \leftarrow \{0_{1,2}, 1_{1,2}\}^n} [\mathsf{Sipser}_d(\mathbf{X}) = 1] = \frac{1}{2} \pm o_n(1). \tag{6}$$

However, since this fact is a direct byproduct of our main theorem (which shows that Sipser_d cannot be $(1/2-o_n(1))$-approximated by any depth-$(d-1)$ formula, let alone by a constant function), we omit the tedious direct analysis here.

We specify an addressing scheme for the gates and input variables of our Sipser_d formula which will be heavily used throughout the paper. Let $A_0 = \{\mathsf{output}\}$, and for $1 \leq k \leq d$, let $A_k = A_{k-1} \times [w_{k-1}]$. An element of A_k specifies the address of a gate at depth (distance from the output node) k in Sipser_d in the obvious way; so $A_d = \{\mathsf{output}\} \times [w_0] \times \cdots \times [w_{d-1}]$ is the set of addresses of the input variables and $|A_d| = n$.

We close this section by introducing notation for the following family of formulas related to Sipser_d:

Definition 5. *For $1 \leq k \leq d$, we write $\mathsf{Sipser}_d^{(k)} : \{0,1\}^{A_k} \to \{0,1\}$ to denote the depth-k formula obtained from Sipser_d by discarding all gates at depths $k+1$ through $d-1$, and replacing every depth-k gate at address $a \in A_k$ with a fresh formal variable y_a.*

Note that $\mathsf{Sipser}_d^{(1)}$ is the top gate of Sipser_d; in particular, $\mathsf{Sipser}_d^{(1)}$ is an w_0-way OR if d is even, and an w_0-way AND if d is odd. Note also that $\mathsf{Sipser}_d^{(d)}$ is simply Sipser_d itself, although we stress that $\mathsf{Sipser}_d^{(k)}$ is not the same as Sipser_k for $1 \leq k \leq d-1$.

VII. Setup for and overview of our proof

A. Key parameter settings

The starting point for our parameter settings is the pair of fixed values

$$\lambda := \frac{(\log w)^{3/2}}{w^{5/4}} \quad \text{and} \quad q := \sqrt{p} = 2^{-m/2}. \tag{7}$$

Given these fixed values of λ and q, we define a sequence of parameters $t_{d-1}, \ldots, t_1$ as

$$t_{d-1} := \frac{p-\lambda}{q}, \qquad t_{k-1} := \frac{(1-t_k)^{qw} - \lambda}{q} \quad \text{for } k = d-1, \ldots, 2. \tag{8}$$

Each of our $d-1$ random projections will be defined with respect to an underlying product distribution. Our first random projection $\mathrm{proj}_{\boldsymbol{\rho}^{(d)}}$ will be associated with the uniform distribution over $\{0,1\}^n$; this is because our ultimate goal is to establish uniform-distribution correlation bounds. For $k \in \{2, \ldots, d-1\}$ the subsequent random projections $\mathrm{proj}_{\boldsymbol{\rho}^{(k)}}$ will be associated with either the t_k-biased or $(1-t_k)$-biased product distribution (depending on whether $d-k$ is even or odd). Recalling our discussion in Section IV of the framework for proving correlation bounds — in particular, the three key properties our random projections have to satisfy — the values for $t_1, \ldots, t_{d-1}$ are chosen carefully so that the compositions of our $d-1$ random projections complete to the uniform distribution, satisfying Property 3 (we prove this in Section VIII of the full version).

The next lemma gives bounds on $t_{d-1}, \ldots, t_1$ which show that these values "stay under control". By our definitions of λ, p and q in (7), we have that $t_{d-1} = q - o(q)$, and we will need the fact that the values of t_k for $k = d-1, \ldots, 2$ remain in the range $q \pm o(q)$. Roughly speaking, since each t_{k-1} is defined inductively in terms of t_k from $k = d-1$ down to 1, we have to argue that these values do not "drift" significantly from the initial value of $t_{d-1} = q - o(q)$. We need to keep these values under control for two reasons: first, the magnitude of these values directly affects the strength of our Projection Switching Lemma — as we will see in Section IX-A of the full version, our error bounds depend on the magnitude of these t_k's. Second, since the top fan-in w_0 of our Sipser_d function is directly determined by t_1 (recall (4)), we need a bound on t_1 to control the structure of this function.

Lemma VII.1. *There is a universal constant $c > 0$ such that for $2 \leq d \leq \frac{cm}{\log m}$, we have that $t_k = q \pm q^{1.1}$ for all $k \in [d-1]$.*

We defer the proof of Lemma VII.1 to the full version. The $k = 1$ case of Lemma VII.1 along with our definition of w_0 (recall (4)) give us the bounds

$$\frac{1}{2} \geq (1-t_1)^{qw_0} \geq \frac{1}{2}(1-tq) = \frac{1}{2}\left(1 - \frac{\Theta(\log w)}{w}\right) = \frac{1}{2}\left(1 - \Theta(2^{-m})\right). \tag{9}$$

These bounds (showing that $(1-t_1)^{qw_0}$ is very close to $1/2$) will be useful for our proof that Sipser_d remains essentially unbiased (i.e. it remains "structured") under our random projections, which in turn implies our claim (6) that Sipser_d is essentially balanced.

We close this subsection with the following estimates of our key parameters in terms of w:

$$p = \Theta\left(\frac{\log w}{w}\right), \quad q = \Theta\left(\sqrt{\frac{\log w}{w}}\right), \quad t_k = \Theta\left(\sqrt{\frac{\log w}{w}}\right) \quad \text{for all } k \in [d-1]. \tag{10}$$

B. The initial and subsequent random projections

As described in Section IV, our overall approach is structured around a sequence of random projections which we will apply to both the target function Sipser_d and the approximating circuit C. Both are functions over $\{0,1\}^n \equiv \{0,1\}^{A_d}$, and our $d-1$ random projections will sequentially transform them from being over $\{0,1\}^{A_k}$ to being over $\{0,1\}^{A_{k-1}}$ for $k = d$ down to $k = 1$. Thus, at the end of the overall process both the randomly projected target and the randomly projected approximator are functions over $\{0,1\}^{A_1} \equiv \{0,1\}^{w_0}$.

We now formally define this sequence of random projections; recalling Definition 4, to define a random projection operator it suffices to specify a distribution over random restrictions, and this is what we will do. We begin with the initial random projection:

Definition 6 (Initial random projection)**.** *The distribution $\mathcal{R}_{\text{init}}$ over restrictions ρ in $\{0,1,*\}^{A_{d-1}\times[m]} \equiv \{0,1,*\}^n$ (recall that $w_{d-1} = m$) is defined as follows: independently for each $a \in A_{d-1}$,*

$$\boldsymbol{\rho}_b \leftarrow \begin{cases} \{1\}^m & \textit{with probability } \lambda \\ \{*_{1/2}, 1_{1/2}\}^m \setminus \{1\}^m & \textit{with probability } q \\ \{0_{1/2}, 1_{1/2}\}^m \setminus \{1\}^m & \textit{with probability } 1-\lambda-q. \end{cases} \tag{11}$$

Remark 10. The description of $\mathcal{R}_{\text{init}}$ given in Definition 6 will be most convenient for our arguments, but we note here the following equivalent view of an $\mathcal{R}_{\text{init}}$-random projection. Let $\mathcal{R}'_{\text{init}}$ be the distribution over restrictions ρ' in $\{0,1,*\}^{A_{d-1}\times[m]} \equiv \{0,1,*\}^n$ where

$$\boldsymbol{\rho}'_a \leftarrow \{*_{1/2}, 1_{1/2}\}^m \setminus \{1\}^m \quad \text{independently for each } a \in A_{d-1},$$

and $\mathcal{R}''_{\text{init}}$ be the distribution of restrictions ρ'' in $\{0,1,*\}^{A_{d-1}}$ where

$$\boldsymbol{\rho}''_a \leftarrow \begin{cases} 1 & \text{with probability } \lambda \\ * & \text{with probability } q \\ 0 & \text{with probability } 1-\lambda-q \end{cases} \qquad \text{independently for each } a \in A_{d-1}.$$

Then for all $f : \{0,1\}^n \to \{0,1\}$ we have that $\mathrm{proj}_{\boldsymbol{\rho}} f$, where $\boldsymbol{\rho} \leftarrow \mathcal{R}_{\text{init}}$, is distributed identically to

$$(\mathrm{proj}_{\boldsymbol{\rho}'} f) \restriction \boldsymbol{\rho}'' \quad \text{where } \boldsymbol{\rho}' \leftarrow \mathcal{R}'_{\text{init}} \text{ and } \boldsymbol{\rho}'' \leftarrow \mathcal{R}''_{\text{init}}.$$

1) Subsequent random projections: Our subsequent random projections will alternate between two types, depending on whether $d-k$ is even or odd. These types are dual to each other in the sense that their distributions are completely identical, except with the roles of 1 and 0 swapped; in other words, the bitwise complement of a draw from the first type yields a draw from the second type. To avoid redundancy in our definitions we introduce the notation in Table II: we represent $\{0,1\}^{A_k}$ as $\{\bullet,\circ\}^{A_k}$, where a $\circ$-value corresponds to either 1 or 0 depending on whether $d-k$ is even or odd, and the $\bullet$-value is simply the complement of the $\circ$-value. For example, the string $(\circ,\circ,\bullet,\circ)$ translates to $(1,1,0,1)$ if $d-k$ is even, and $(0,0,1,0)$ if $d-k$ is odd.

	Gates of Sipser_d at depth $k-1$	$\circ$	$\bullet$
$d-k \equiv 0 \mod 2$	AND	1	0
$d-k \equiv 1 \mod 2$	OR	0	1

Table II: Conversion table for $\tau \in \{\bullet,\circ,*\}^{A_k}$ where $1 \le k \le d$.

In an interesting contrast with Håstad's proofs of the worst-case depth hierarchy theorem (Theorem 4) and of $\mathsf{Parity} \notin \mathsf{AC}^0$, our stage-wise random projection process is *adaptive*: apart from the initial $\mathcal{R}_{\text{init}}$-random projection, the distribution of each random projection depends on the outcome of the previous. We will need the following notion of the "lift" of a restriction to describe this dependence:

Definition 7 (Lift)**.** *Let $2 \le k \le d$ and $\tau \in \{\bullet, \circ, *\}^{A_{k-1} \times [w_{k-1}]} \equiv \{\bullet, \circ, *\}^{A_k}$. The* lift of τ *is the string $\widehat{\tau} \in \{\bullet, \circ, *\}^{A_{k-1}}$ defined as follows: for each $a \in A_{k-1}$, the coordinate $\widehat{\tau}_a$ of $\widehat{\tau}$ is*

$$\widehat{\tau}_a = \begin{cases} \circ & \text{if } \tau_{a,i} = \bullet \text{ for any } i \in [w_{k-1}] \\ \bullet & \text{if } \tau_a = \{\circ\}^{w_{k-1}} \\ * & \text{if } \tau_a \in \{*, \circ\}^{w_{k-1}} \setminus \{\circ\}^{w_{k-1}}. \end{cases}$$

We remind the reader that $\tau \in \{\bullet, \circ, *\}^{A_k}$ and $\widehat{\tau} \in \{\bullet, \circ, *\}^{A_{k-1}}$ belong to adjacent levels (i.e. they fall under different rows in Table II). Consequently, for example, if 1 corresponds to $\bullet$ as a symbol in τ then it corresponds to $\circ$ as a symbol in $\widehat{\tau}$, and vice versa.

Later this notion of the "lift" of a restriction will also be handy when we describe the effect of our random projections on the target function Sipser_d. The high-level rationale behind it is that $\widehat{\tau} \in \{\bullet, \circ, *\}^{A_{k-1}}$ denotes the values that the bottom-layer gates of $\mathsf{Sipser}_d^{(k)}$ take on when its input variables are set according to $\tau \in \{\bullet, \circ, *\}^{A_k}$. As a concrete example, suppose $d - k \equiv 0 \mod 2$ and let $\tau \in \{0, 1, *\}^{A_k}$ be a restriction. Since $d - k \equiv 0 \mod 2$, recalling Table II we have that the bottom-layer gates of $\mathsf{Sipser}_d^{(k)}$ (or equivalently, the gates of Sipser_d at depth $k-1$) are AND gates. For every block $a \in A_{k-1}$,

- If $\tau_{a,i} = 0$ for some $i \in [w_{k-1}]$, the AND gate at address a is falsified and has value 0.
- If $\tau_{a,i} = \{1\}^{w_{k-1}}$, the AND gate at address a is satisfied and has value 1.
- If $\tau_a \in \{*, 1\} \setminus \{1\}^{w_{k-1}}$, the value of the AND gate at address a remains undetermined (which we denote as having value $*$).

These three cases correspond exactly to the three branches in Definition 7, and so indeed $\widehat{\tau}_a \in \{0, 1, *\}$ represents the value that the AND gate at address a takes when its input variables are set according to $\tau_a \in \{0, 1, *\}^{w_{k-1}}$.

We shall require the following technical definition:

Definition 8 (k-acceptable)**.** *For $2 \le k \le d-1$ and a set $S \subseteq [w_{k-1}]$, we say that S is* k-acceptable *if*

$$|S| = qw \pm w^{\beta(k,d)}, \quad \text{where } \beta(k,d) := \frac{1}{3} + \frac{d-k-1}{12d}.$$

Note that $\frac{1}{3} \le \beta(k,d) \le \frac{5}{12} < \frac{1}{2}$ for all $d \in \mathbb{N}$ and $2 \le k \le d-1$.

For intuition, in the above definition S should be thought of as specifying those children of a particular depth-$(k-1)$ gate of Sipser_d that take the value $*$ under certain restrictions (defined below). We want the size of this set to be essentially qw, and as k gets smaller (closer to the root), for technical reasons we allow more and more — but never too much — deviation from this desired value. See Section X-A of the full version for a detailed discussion.

We are now ready to give the key definition for our subsequent random projections:

Definition 9 (Subsequent random projections)**.** *Let $\tau \in \{\bullet, \circ, *\}^{A_k}$ where $2 \le k \le d-1$. We define a distribution $\mathcal{R}(\tau)$ over refinements $\boldsymbol{\rho} \in \{\bullet, \circ, *\}^{A_k}$ of τ as follows. Independently for each $a \in A_{k-1}$, writing $S_a = S_a(\tau)$ to denote $\tau_a^{-1}(*) = \{i \in [w_{k-1}] \colon \tau_{a,i} = *\}$ and $\boldsymbol{\rho}(S_a)$ to denote the substring of $\boldsymbol{\rho}_a$ with coordinates in S_a,*

- *If $\widehat{\tau}_a = \circ$ (i.e. if $\tau_{a,i} = \bullet$ for some $i \in [w_{k-1}]$) or if S_a is not k-acceptable, then*

$$\boldsymbol{\rho}(S_a) \leftarrow \{\bullet_{t_k}, \circ_{1-t_k}\}^{S_a}.$$

- *If $\widehat{\tau}_a = *$ (i.e. if $\tau_{a,i} \in \{*, \circ\}^{w_{k-1}} \setminus \{\circ\}^{w_{k-1}}$) and S_a is k-acceptable, then*

$$\boldsymbol{\rho}(S_a) \leftarrow \begin{cases} \circ^{S_a} & \text{with probability } \lambda \\ \{*_{t_k}, \circ_{1-t_k}\}^{S_a} \setminus \{\circ\}^{S_a} & \text{with probability } q_a \\ \{\bullet_{t_k}, \circ_{1-t_k}\}^{S_a} \setminus \{\circ\}^{S_a} & \text{with probability } 1 - \lambda - q_a, \end{cases} \tag{12}$$

where

$$q_a := \frac{(1-t_k)^{|S_a|} - \lambda}{t_{k-1}} \quad \text{is chosen to satisfy } (1-t_k)^{|S_a|} = \lambda + q_a t_{k-1}. \tag{13}$$

(Note that if $\widehat{\tau}_a = \bullet$ then $\tau_{a,i} = \circ$ for all $i \in [w_{k-1}]$, and so τ_a cannot be refined further.)

For all $a \in A_{k-1}$ and $i \in [w_{k-1}]$ such that $\tau_{a,i} \in \{\bullet, \circ\}$, we set $\boldsymbol{\rho}_{a,i} = \tau_{a,i}$ and so $\boldsymbol{\rho}$ is indeed a refinement of τ.

Remark 11. We remark that q_a as defined in (13) is indeed a well-defined quantity in $[0,1]$ if S_a is k-acceptable. We omit the straightforward verification here since our analysis in Section X-A of the full version will in fact establish a stronger statement showing that $q_a = q \pm o(q)$.

Remark 12. By inspecting Definition 6, we see that for all $\rho \in \mathrm{supp}(\mathcal{R}_{\mathrm{init}})$ and blocks $a \in A_{d-1}$

$$\begin{aligned} \rho_{a,i} = * \text{ for some } i \in [m] \quad &\text{iff} \quad \rho_a \in \{*,1\}^m \setminus \{1\}^m, \quad \text{or equivalently,} \\ \rho_{a,i} = * \text{ for some } i \in [m] \quad &\text{iff} \quad \widehat{\rho}_a = *, \end{aligned}$$

and hence for all $h : \{0,1\}^n \to \{0,1\}$ the projection $\mathrm{proj}_\rho h : \{0,1\}^{A_{d-1}} \to \{0,1\}$ depends only on the coordinates in $(\widehat{\rho})^{-1}(*) \subseteq A_{d-1}$. Likewise, by inspecting Definition 9 we have that for all $\tau \in \{\bullet, \circ, *\}^{A_k}, \rho \in \mathrm{supp}(\mathcal{R}(\tau))$, and blocks $a \in A_{k-1}$,

$$\begin{aligned} \rho_{a,i} = * \text{ for some } i \in [w_{k-1}] \quad &\text{iff} \quad \rho_a \in \{*,\circ\}^{w_{k-1}} \setminus \{\circ\}^{w_{k-1}}, \quad \text{or equivalently,} \\ \rho_{a,i} = * \text{ for some } i \in [w_{k-1}] \quad &\text{iff} \quad \widehat{\rho}_a = *, \end{aligned}$$

and hence for all $h : \{0,1\}^{A_k} \to \{0,1\}$ the projection $\mathrm{proj}_\rho h : \{0,1\}^{A_{k-1}} \to \{0,1\}$ depends only on the coordinates in $(\widehat{\rho})^{-1}(*) \subseteq A_{k-1}$. Our proof that our sequence of random projections (based on Definitions 6 and 9 as described in Definition 4) completes to the uniform distribution will rely on these properties; see Section VIII of the full version.

C. Overview of our proof

With the definitions from Section VII-B in hand, we are (finally) in a position to give a detailed overview of our proof. Let C be a depth-d approximating circuit for Sipser_d, where C either has significantly smaller bottom fan-in than Sipser_d (in the case of Theorem 6) or the opposite alternation pattern to Sipser_d (in the case of Theorem 7), and C satisfies the size bounds given in the respective theorem statements. In both cases our goal is to show that C has small correlation with Sipser_d, i.e. to prove that

$$\mathbf{Pr}[\mathsf{Sipser}_d(\mathbf{X}) \neq C(\mathbf{X})] \geq \frac{1}{2} - o_n(1) \tag{14}$$

for a uniform random input $\mathbf{X} \leftarrow \{0_{1/2}, 1_{1/2}\}^n$. At a high level, we do this by analyzing the effect of $d-1$ random projections on the target and the approximator: we begin with an $\mathcal{R}_{\mathrm{init}}$-random projection $\mathrm{proj}_{\boldsymbol{\rho}^{(d)}}$ where $\boldsymbol{\rho}^{(d)} \leftarrow \mathcal{R}_{\mathrm{init}}$, followed by $\mathrm{proj}_{\boldsymbol{\rho}^{(d-1)}}$ where $\boldsymbol{\rho}^{(d-1)} \leftarrow \mathcal{R}(\widehat{\boldsymbol{\rho}^{(d)}})$, and then $\mathrm{proj}_{\boldsymbol{\rho}^{(d-2)}}$ where $\boldsymbol{\rho}^{(d-2)} \leftarrow \mathcal{R}(\widehat{\boldsymbol{\rho}^{(d-1)}})$, and so on. It is interesting to note that unlike Håstad's proofs of the worst-case depth hierarchy theorem (Theorem 4) and of $\mathsf{Parity} \notin \mathsf{AC}^0$, the distribution of our k-th random projection is defined *adaptively* depending on the outcome of the $(k-1)$-st. For notational concision we introduce the following definition for this overall $(d-1)$-stage projection:

Definition 10. *Given a function $f : \{0,1\}^n \to \{0,1\}$, we write $\boldsymbol{\Psi}(f) : \{0,1\}^{w_0} \to \{0,1\}$ to denote the following random projection of f:*

$$\boldsymbol{\Psi}(f) \equiv \mathrm{proj}_{\boldsymbol{\rho}^{(2)}} \mathrm{proj}_{\boldsymbol{\rho}^{(3)}} \cdots \mathrm{proj}_{\boldsymbol{\rho}^{(d-1)}} \mathrm{proj}_{\boldsymbol{\rho}^{(d)}} f,$$

where $\boldsymbol{\rho}^{(d)} \leftarrow \mathcal{R}_{\mathrm{init}}$ and $\boldsymbol{\rho}^{(k)} \leftarrow \mathcal{R}(\widehat{\boldsymbol{\rho}^{(k+1)}})$ for all $2 \leq k \leq d-1$. We will sometimes refer to the overall process as a $\boldsymbol{\Psi}$-random projection, and $\boldsymbol{\Psi}(f)$ as the $\boldsymbol{\Psi}$-random projection of f. (We remind the reader that the projection of a function over $\{0,1\}^{A_k}$ yields a function over $\{0,1\}^{A_{k-1}}$ for all $2 \leq k \leq d$, and in particular $\boldsymbol{\Psi}(f)$ is indeed a function over $\{0,1\}^{A_1} \equiv \{0,1\}^{w_0}$.)

Recalling the framework for proving correlation bounds discussed in Section IV, the rest of the paper is structured around showing that a $\boldsymbol{\Psi}$-random projection satisfies the three key properties outlined in Section IV:

Property 1. The approximating circuit C simplifies under a $\boldsymbol{\Psi}$-random projection.
Property 2. The target Sipser_d remains structured under a $\boldsymbol{\Psi}$-random projection.
Property 3. $\boldsymbol{\Psi}$ completes to the uniform distribution.

Section VIII (of the full version): We begin in Section VIII with Property 3. We show that

$$\mathbf{Pr}[\mathsf{Sipser}_d(\mathbf{X}) \neq C(\mathbf{X})] = \mathbf{Pr}[(\mathbf{\Psi}(\mathsf{Sipser}_d))(\mathbf{Y}) \neq (\mathbf{\Psi}(C))(\mathbf{Y})] \tag{15}$$

where $\mathbf{Y}$ is drawn from an appropriate product distribution $\mathcal{D}$ over $\{0,1\}^{w_0}$ ($\mathcal{D}$ is the t_1-biased product distribution if d is even, and $(1-t_1)$-biased product distribution if d is odd). This reduces our goal of bounding the correlation between Sipser_d and C (i.e. (14)) under the uniform distribution, to the task of bounding the correlation between their $\mathbf{\Psi}$-random projections $\mathbf{\Psi}(\mathsf{Sipser}_d)$ and $\mathbf{\Psi}(C)$ with respect to $\mathcal{D}$.

Section IX (of the full version): With the reduction (15) in hand, we turn our attention to Property 1, showing that the approximating circuit C of the type specified in either Theorems 6 or 7 "collapses to a simple function" under a $\mathbf{\Psi}$-random projection. More precisely, for the case that the depth-d circuit C has significantly smaller bottom fan-in than Sipser_d we show that C collapses to a shallow decision tree, and for the case that C has the opposite alternation pattern to Sipser_d we show that C collapses to a small-width depth-two circuit with top gate opposite to that of $\mathbf{\Psi}(\mathsf{Sipser}_d)$. (In both cases these statements are with high probability under a $\mathbf{\Psi}$-random projection.)

In close parallel with Håstad's "bottom-up" proof of $\mathsf{Parity} \notin \mathsf{AC}^0$, the main technical ingredient in this section is a *projection switching lemma* showing that the random projection $\mathrm{proj}_{\boldsymbol{\rho}^{(k)}}$ of a small-width DNF or CNF "switches" to a small-depth decision tree with high probability. Applying this lemma to every bottom-level depth-2 subcircuit of C, we are able to argue that each of the $d-1$ random projections comprising $\mathbf{\Psi}$ reduces the depth of C by one with high probability, and thus $\mathbf{\Psi}(C)$ collapses to a small-depth decision tree or small-width depth-two circuit as claimed.

Section X (of the full version): It remains to argue that the target Sipser_d — in contrast with the approximating circuit C — "retains structure" with high probability under a $\mathbf{\Psi}$-random restriction. This is a high-probability statement because there is a nonzero failure probability introduced by each of the $d-1$ individual random projections $\mathrm{proj}_{\boldsymbol{\rho}^{(k)}}$ that comprise $\mathbf{\Psi} \equiv \{\boldsymbol{\rho}^{(k)}\}_{k \in \{2,\ldots,d\}}$. To reason about and bound these failure probabilities we introduce the notion of a "typical" restriction. The parameters of our definition of typicality are chosen carefully to ensure that

(i) $\boldsymbol{\rho}^{(d)} \leftarrow \mathcal{R}_{\mathrm{init}}$ is typical with high probability, and
(ii) if $\rho^{(k+1)}$ is typical, then $\boldsymbol{\rho}^{(k)} \leftarrow \mathcal{R}(\widehat{\rho^{(k+1)}})$ is also typical with high probability.

We establish (i) and (ii) in Section X-A. Together, (i) and (ii) imply that with high probability $\mathbf{\Psi} \equiv \{\boldsymbol{\rho}^{(k)}\}_{k \in \{2,\ldots,d\}}$ is such that $\boldsymbol{\rho}^{(d)}, \ldots, \boldsymbol{\rho}^{(2)}$ are *all* typical; we use this in Section X-B.

With the notion of typical restrictions in hand, in Section X-B we establish Property 2 showing that Sipser_d "survives" a $\mathbf{\Psi}$-random projection (i.e. it "retains structure") with high probability. More formally, for outcomes $\Psi \equiv \{\rho^{(k)}\}_{k \in \{2,\ldots,d\}}$ of $\mathbf{\Psi}$ such that $\rho^{(d)}, \ldots, \rho^{(2)}$ are all typical, we prove that the Ψ-projected target $\Psi(\mathsf{Sipser}_d)$ is "well-structured" in the following sense:

(i) $\Psi(\mathsf{Sipser}_d)$ is a depth-one formula: an OR if d is even, an AND if d is odd.
(ii) The bias of $\Psi(\mathsf{Sipser}_d)$ under $\mathcal{D}$ is close to $1/2$; that is,

$$\mathrm{bias}(\Psi(\mathsf{Sipser}_d), \mathbf{Y}) = \frac{1}{2} - o_n(1).$$

Recall that we have shown in Section X-A that with high probability $\mathbf{\Psi} \equiv \{\boldsymbol{\rho}^{(k)}\}_{k \in \{2,\ldots,d\}}$ is such that $\boldsymbol{\rho}^{(d)}, \ldots, \boldsymbol{\rho}^{(2)}$ are all typical. Therefore, the results of these two subsections together imply that the randomly projected target $\mathbf{\Psi}(\mathsf{Sipser}_d)$ satisfies both (i) and (ii) with high probability.

Section XI (of the full version): Having established Properties 1, 2, and 3, it remains to bound the correlation between a depth-one formula with bias essentially $1/2$ and a small-width CNF formula of opposite alternation with respect to the product distribution $\mathcal{D}$ over $\{0,1\}^{w_0}$. (Recall that our results from Section X-B show that $\mathbf{\Psi}(\mathsf{Sipser}_d)$ collapses to the former with high probability, and our results from Section IX shows that $\mathbf{\Psi}(C)$ collapses to the latter with high probability — this holds in both cases since a shallow decision tree is a small-width CNF.) We prove this correlation bound using a slight extension of an argument in [OW07], and with this final piece in hand our main theorems follow from straightforward arguments putting the pieces together.

REFERENCES

[Aar] Scott Aaronson. The Complexity Zoo. Available at http://cse.unl.edu/~cbourke/latex/ComplexityZoo.pdf.

[Aar10a] Scott Aaronson. A counterexample to the generalized Linial-Nisan conjecture. *Electronic Colloquium on Computational Complexity*, 17:109, 2010.

[Aar10b] Scott Aaronson. BQP and the polynomial hierarchy. In *Proceedings of the 42nd ACM Symposium on Theory of Computing*, pages 141–150, 2010.

[Ajt83] Miklós Ajtai. Σ_1^1-formulae on finite structures. *Annals of Pure and Applied Logic*, 24(1):1–48, 1983.

[Ajt94] Miklós Ajtai. The independence of the modulo p counting principles. In *Proceedings of the 26th Annual ACM Symposium on Theory of Computing*, pages 402–411, 1994.

[AWY15] Amir Abboud, Ryan Williams, and Huacheng Yu. More applications of the polynomial method to algorithm design. In *Proceedings of the 26th Annual ACM-SIAM Symposium on Discrete Algorithms*, 2015.

[Bab87] László Babai. Random oracles separate PSPACE from the polynomial-time hierarchy. *Information Processing Letters*, 26(1):51–53, 1987.

[Baz09] Louay Bazzi. Polylogarithmic independence can fool DNF formulas. *SIAM Journal on Computing*, 38(6):2220–2272, 2009.

[BG81] Charles Bennett and John Gill. Relative to a random oracle A, $\mathsf{P}^A \neq \mathsf{NP}^A \neq \mathsf{coNP}^A$ with probability 1. *SIAM Journal on Computing*, 10(1):96–113, 1981.

[BGS75] Theodore Baker, John Gill, and Robert Solovay. Relativizations of the P=?NP question. *SIAM Journal on computing*, 4(4):431–442, 1975.

[BIS12] Paul Beame, Russell Impagliazzo, and Srikanth Srinivasan. Approximating AC^0 by small height decision trees and a deterministic algorithm for $\#\mathsf{AC}^0$-SAT. In *Proceedings of the 27th Conference on Computational Complexity*, pages 117–125, 2012.

[BKS99] Itai Benjamini, Gil Kalai, and Oded Schramm. Noise sensitivity of Boolean functions and applications to percolation. *Inst. Hautes Études Sci. Publ. Math.*, 90:5–43, 1999.

[Boo94] Ronald Book. On collapsing the polynomial-time hierarchy. *Information Processing Letters*, 52(5):235–237, 1994.

[Bop97] Ravi Boppana. The average sensitivity of bounded-depth circuits. *Information Processing Letters*, 63(5):257–261, 1997.

[Bra10] Mark Braverman. Polylogarithmic independence fools AC^0 circuits. *Journal of the ACM*, 57(5):28, 2010.

[BS79] Theodore Baker and Alan Selman. A second step toward the polynomial hierarchy. *Theoretical Computer Science*, 8(2):177–187, 1979.

[BT96] Nader Bshouty and Christino Tamon. On the Fourier spectrum of monotone functions. *Journal of the ACM*, 43(4):747–770, 1996.

[Cai86] Jin-Yi Cai. With probability one, a random oracle separates PSPACE from the polynomial-time hierarchy. In *Proceedings of the 18th Annual ACM Symposium on Theory of Computing*, pages 21–29, 1986.

[CCH98] Liming Cai, Jianer Chen, and Johan Håstad. Circuit bottom fan-in and computational power. *SIAM Journal on Computing*, 27(2):341–355, 1998.

[DK00] Ding-Zhu Du and Ker-I Ko. *Theory of Computational Complexity*. John Wiley & Sons, Inc., 2000.

[For99] Lance Fortnow. Relativized worlds with an infinite hierarchy. *Information Processing Letters*, 69(6):309–313, 1999.

[FSS81] Merrick Furst, James Saxe, and Michael Sipser. Parity, circuits, and the polynomial-time hierarchy. In *Proceedings of the 22nd IEEE Annual Symposium on Foundations of Computer Science*, pages 260–270, 1981.

[GW13] Oded Goldreich and Avi Wigderson. On the size of depth-three Boolean circuits for computing multilinear functions. *Electronic Colloquium on Computational Complexity*, 2013.

[Hås86a] Johan Håstad. Almost optimal lower bounds for small depth circuits. In *Proceedings of the 18th Annual ACM Symposium on Theory of Computing*, pages 6–20, 1986.

[Hås86b] Johan Håstad. *Computational Limitations for Small Depth Circuits*. MIT Press, Cambridge, MA, 1986.

[Hås89] Johan Håstad. *Almost optimal lower bounds for small depth circuits*, pages 143–170. Advances in Computing Research, Vol. 5. JAI Press, 1989.

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

[Hat14] Hamed Hatami. Scribe notes for the course *COMP760: Harmonic Analysis of Boolean Functions*, 2014. Available at http://cs.mcgill.ca/~hatami/comp760-2014/lectures.pdf.

[Hem94] Lane Hemaspaandra. Complexity theory column 5: the not-ready-for-prime-time conjectures. *ACM SIGACT News*, 25(2):5–10, 1994.

[HMP+93] András Hajnal, Wolfgang Maass, Pavel Pudlák, Mário Szegedy, and György Turán. Threshold circuits of bounded depth. *Journal of Computer and System Sciences*, 46:129–154, 1993.

[HO02] Lane Hemaspaandra and Mitsunori Ogihara. *The Complexity Theory Companion*. Springer, 2002.

[HRZ95] Lane Hemaspaandra, Ajit Ramachandran, and Marius Zimand. Worlds to die for. *ACM SIGACT News*, 26(4):5–15, 1995.

[IMP12] Russell Impagliazzo, William Matthews, and Ramamohan Paturi. A satisfiability algorithm for AC^0. In *Proceedings of the 23rd Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 961–972, 2012.

[IS01] Russell Impagliazzo and Nathan Segerlind. Counting axioms do not polynomially simulate counting gates. In *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science*, pages 200–209, 2001.

[Joh86] David Johnson. The NP-completeness column: An ongoing guide. *Journal of Algorithms*, 7(2):289–305, 1986.

[Juk12] Stasys Jukna. *Boolean Function Complexity*. Springer, 2012.

[Kal00] Gil Kalai. Combinatorics with a geometric flavor: some examples, 2000. GAFA Special Volume 10, Birkhauser Verlag, Basel, 2000.

[Kal10] Gil Kalai. Noise Stability and Threshold Circuits. Blog post at *Combinatorics and more*, 2010. https://gilkalai.wordpress.com/2010/02/10/noise-stability-and-threshold-circuits.

[Kal12] Gil Kalai. Answer to the question: *Are all functions whose Fourier weight is concentrated on the small sized sets computed by* AC^0 *circuits?* Theoretical Computer Science StackExchange, 2012. http://cstheory.stackexchange.com/questions/12769/are-all-the-functions-whose-fourier-weight-is-concentrated-on-the-small -sized-se.

[KPPY84] Maria Klawe, Wolfgang Paul, Nicholas Pippenger, and Mihalis Yannakakis. On monotone formulae with restricted depth. In *Proceedings of the 16th Annual ACM Symposium on Theory of Computing*, pages 480–487, 1984.

[KPW95] Jan Krajíček, Pavel Pudlák, and Alan Woods. An exponential lower bound to the size of bounded depth frege proofs of the pigeonhole principle. *Random Structures & Algorithms*, 7(1):15–39, 1995.

[KS05] Gil Kalai and Shmuel Safra. Threshold phenomena and influence. In *Computational Complexity and Statistical Physics*, pages 25–60. Oxford University Press, 2005.

[LMN93] Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, fourier transform, and learnability. *Journal of the ACM*, 40(3):607–620, 1993.

[Man95] Yishay Mansour. An $O(n^{\log\log n})$ learning algorithm for DNF under the uniform distribution. *Journal of Computer and System Sciences*, 50:543–550, 1995.

[Nis91] Noam Nisan. Pseudorandom bits for constant depth circuits. *Combinatorica*, 11(1):63–70, 1991.

[O'D07] Ryan O'Donnell. Lecture 29: Open Problems. Scribe notes for the course *CMU 18-859S: Analysis of Boolean Functions*, 2007. Available at http://www.cs.cmu.edu/~odonnell/boolean-analysis.

[OW07] Ryan O'Donnell and Karl Wimmer. Approximation by DNF: examples and counterexamples. In *34th International Colloquium on Automata, Languages and Programming*, pages 195–206, 2007.

[PBI93] Toniann Pitassi, Paul Beame, and Russell Impagliazzo. Exponential lower bounds for the pigeonhole principle. *Computational complexity*, 3(2):97–140, 1993.

[Raz87] Alexander Razborov. Lower bounds on the size of bounded depth circuits over a complete basis with logical addition. *Mathematical Notes of the Academy of Sciences of the USSR*, 41(4):333–338, 1987.

[Raz09] Alexander Razborov. A simple proof of Bazzi's theorem. *ACM Transactions on Computation Theory*, 1(1):3, 2009.

[SBI04] Nathan Segerlind, Sam Buss, and Russell Impagliazzo. A switching lemma for small restrictions and lower bounds for k-DNF resolution. *SIAM Journal on Computing*, 33(5):1171–1200, 2004.

[Sip83] Michael Sipser. Borel sets and circuit complexity. In *Proceedings of the 15th Annual ACM Symposium on Theory of Computing*, pages 61–69, 1983.

[Smo87] Roman Smolensky. Algebraic methods in the theory of lower bounds for boolean circuit complexity. In *Proceedings of the 19th Annual ACM Symposium on Theory of Computing*, pages 77–82, 1987.

[ST95] David Shmoys and Éva Tardos. Computational Complexity. In *Handbook of Combinatorics (Ronald Graham, Martin Grötschel, and László Lovász, eds.)*, volume 2. North-Holland, 1995.

[Sub61] Bella Subbotovskaya. Realizations of linear functions by formulas using $\vee$, &, $-$. *Doklady Akademii Nauk SSSR*, 136(3):553–555, 1961.

[Tar89] Gábor Tardos. Query complexity, or why is it difficult to separate $\mathsf{NP}^A \cap \mathsf{coNP}^A$ from P^A by random oracles A? *Combinatorica*, 9(4):385–392, 1989.

[Val83] Leslie Valiant. Exponential lower bounds for restricted monotone circuits. In *Proceedings of the 15th Annual ACM Symposium on Theory of Computing*, pages 110–117, 1983.

[Vio13] Emanuele Viola. Challenges in computational lower bounds. *Electronic Colloquium on Computational Complexity*, 2013.

[VW97] Heribert Vollmer and Klaus Wagner. *Measure One Results in Computational Complexity Theory*, pages 285–312. Advances in Algorithms, Languages, and Complexity. Springer, 1997.

[Wil14a] Ryan Williams. Faster all-pairs shortest paths via circuit complexity. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 664–673, 2014.

[Wil14b] Ryan Williams. The polynomial method in circuit complexity applied to algorithm design (invited survey). In *Proceedings of the 34th Foundations of Software Technology and Theoretical Computer Science Conference*, 2014.

[Yao85] Andrew Yao. Separating the polynomial-time hierarchy by oracles. In *Proceedings of the 26th Annual Symposium on Foundations of Computer Science*, pages 1–10, 1985.