An average-case depth hierarchy theorem for Boolean circuits

Johan Håstad, KTH Stockholm
Benjamin Rossman, NII and Simons Institute, UC Berkeley
Rocco A. Servedio, Columbia University
Li-Yang Tan, Simons Institute, UC Berkeley

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 [Håstad 1986b].

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 [Håstad 1986a], Cai [Cai 1986], and Babai [Babai 1987]. We also use our result to show that there is no "approximate converse" to the results of Linial, Mansour, Nisan [Linial et al. 1993] and Boppana [Boppana 1997] on the total influence of bounded-depth circuits.

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

CCS Concepts: •**Theory of computation** → **Circuit complexity;**

Additional Key Words and Phrases: Boolean circuit complexity, polynomial hierarchy, random oracles, random projections

ACM Reference Format:
Johan Håstad, Benjamin Rossman, Rocco A. Servedio, Li-Yang Tan, 2016. An average-case depth hierarchy theorem for Boolean circuits. *J. ACM* 9, 4, Article 39 (March 2016), 29 pages.
DOI: 0000001.0000001

1. 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 [Yao 1985; Håstad 1986a; Razborov 1987; Smolensky 1987] 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 [Linial et al. 1993; Mansour 1995], pseudorandomness [Nisan 1991; Bazzi 2009; Razborov 2009; Braverman 2010], proof complexity [Pitassi et al. 1993; Ajtai 1994; Krajíček et al. 1995], structural complexity [Yao 1985; Håstad 1986a; Cai 1986], and even algorithm design [Williams 2014a; Williams 2014b; Abboud et al. 2015].

In addition to *worst-case* lower bounds against small-depth circuits, *average-case* lower bounds, or *correlation bounds*, have also received significant attention. As

R.A.S. is supported by NSF grants CCF-1319788 and CCF-1420349. B.R. is supported by NSERC. Part of this research was done while L.-Y.T. was visiting Columbia University.
Author's addresses: J. Håstad, KTH - Royal Institute of Technology, B. Rossman, University of Toronto; R. A. Servedio, Computer Science Department, Columbia University; L.-Y. Tan, Toyota Technological Institute at Chicago

DOI: 0000001.0000001

one recent example, Impagliazzo, Matthews, Paturi [Impagliazzo et al. 2012] and Håstad [Håstad 2014] 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 [Ajtai 1983; Yao 1985; Håstad 1986a; Cai 1986; Babai 1987; Beame et al. 2012]. 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 1.1. *For sufficiently large n and $2 \leq d \leq \frac{c \log n}{\log \log n}$, where $c > 0$ is an absolute constant, there is an n-variable Boolean function Sipser_d that is computed by a read-once monotone depth-d formula and that has the following property: Any circuit C of depth at most $d-1$ and size at most $2^{n^{1/5(d-1)}}$ agrees with Sipser_d on at most $\frac{1}{2}+n^{-\Omega(1/d)}$ fraction of all inputs.*

(We actually prove two incomparable lower bounds, each of which implies Theorem 1.1 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 1.1 is an average-case extension of the worst-case depth hierarchy theorems of Sipser, Yao, and Håstad [Sipser 1983; Yao 1985; Håstad 1986a], and answers an open problem of Håstad [Håstad 1986a] (which also appears in [Håstad 1986b; Håstad 1989]). A version of Theorem 1.1 for depths d up to $O(\sqrt{\log n})$ was obtained in [Rossman et al. 2015b]. The improved parameters of Theorem 1.1 were obtained in subsequent work of [Håstad 2016].

We discuss the background and context for Theorem 1.1 in Section 1.1, and state our two main lower bounds more precisely in Section 1.2.

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 [Furst et al. 1981; Sipser 1983], Theorem 1.1 implies that the polynomial hierarchy is infinite relative to a random oracle:

THEOREM 1.2. *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 [Håstad 1986a; Cai 1986; Babai 1987] and has subsequently been discussed in a wide range of surveys [Johnson 1986; Hemaspaandra 1994; Shmoys and Tardos 1995; Hemaspaandra et al. 1995; Vollmer and Wagner 1997; Aaronson], textbooks [Du and Ko 2000; Hemaspaandra and Ogihara 2002], and research papers [Håstad 1986b; Håstad 1989; Tardos 1989; Fortnow 1999; Aaronson 2010a]. (Indeed, the results of [Håstad 1986a; Cai 1986; Babai 1987], 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 2 for details.

Our second application is a strong negative answer a question of Hatami in the analysis of Boolean functions. Seeking an *approximate converse* to the fundamental results of Linial, Mansour, Nisan [Linial et al. 1993] and Boppana [Boppana 1997] on

the total influence of bounded-depth circuits, Hatami asked whether every Boolean function with total influence $\log(n)$ can be approximated by a constant-depth circuit of polynomial size [Hatami 2014]. Answering this question, as a consequence of Theorem 1.1, we obtain the following:

THEOREM 1.3.

(1) *For every constant d, there is a sequence of monotone functions $f^{(d)} : \{0,1\}^n \to \{0,1\}$ with total influence $\leq \log n$ such that $f^{(d)}$ cannot be approximated on 51% of inputs by depth-d circuits of size* $\exp(\exp(\Omega((\log n)^{1/d})))$.
(2) *There is a sequence of monotone functions $f : \{0,1\}^n \to \{0,1\}$ with total influence $\leq \log n$ such that f cannot be approximated on 51% of inputs by depth-$\sqrt{\log \log n}$ circuits of size* $\exp(\exp(\Omega(2^{\sqrt{\log \log n}})))$.

Theorem 1.3 shows that the total influence bound of [Linial et al. 1993; Boppana 1997] does not admit even a very weak approximate converse in the bounded-depth setting. See Section 3 for details.

1.1. 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 1.2 and 1.3, in Sections 2 and 3 respectively.

Sipser was the first to prove a worst-case depth hierarchy theorem for small-depth circuits [Sipser 1983]. 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 $c_d > 0$ [Yao 1985]. Shortly thereafter Håstad proved a near-optimal separation for (a slight variant of) the Sipser functions:

THEOREM 1.4 ([HÅSTAD 1986A]; SEE ALSO [HÅSTAD 1986B; HÅSTAD 1989]). *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 [Cai et al. 1998], and Segerlind, Buss, and Impagliazzo [Segerlind et al. 2004]. Prior to the work of Yao and Håstad, Klawe, Paul, Pippenger, and Yannakakis [Klawe et al. 1984] 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 [O'Donnell and Wimmer 2007]. 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 1.5 (THEOREM 1.9 OF [O'DONNELL AND WIMMER 2007]). *For $w \in \mathbb{N}$ and $n := w2^w$, let* 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 1.1 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 [O'Donnell and Wimmer 2007] 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 [Jukna 2012] and [Valiant 1983; Goldreich and Wigderson 2013; Viola 2013]). In particular, an extension of the $2^{\Omega(n/\mathrm{polylog}(n))}$-type lower bound of [O'Donnell and Wimmer 2007] to depth 3, even for worst-case computation, would constitute a significant breakthrough.

1.2. 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 1.6 (FIRST MAIN LOWER BOUND). *For n sufficiently large and $2 \leq d \leq \frac{c \log n}{\log \log n}$, 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 $2^{n^{1/5(d-1)}}$ and bottom fan-in $\frac{\log n}{10(d-1)}$ agrees with Sipser_d on at most $\frac{1}{2} + n^{-\Omega(1/d)}$ fraction of inputs.*

THEOREM 1.7 (SECOND MAIN LOWER BOUND). *For n sufficiently large and $2 \leq d \leq \frac{c \log n}{\log \log n}$, 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 $2^{n^{1/5(d-1)}}$ 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 $\frac{1}{2} + n^{-\Omega(1/d)}$ fraction of inputs.*

Clearly both these results imply Theorem 1.1 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 1.6 and 1.7. In fact, Theorems 1.6 and 1.7 both follow from an even stronger result, Theorem 10.2, showing that Sipser_d cannot be approximated by depth $d+1$ circuits with restricted bottom fan-in and the opposite alternation pattern.

(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 [Bshouty and Tamon 1996] 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. [Hajnal et al. 1993] 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 1.6 or Theorem 1.7 "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 4 for a much more detailed explanation of our techniques (and of why previous approaches were unable to successfully establish the result).

2. APPLICATION #1: RANDOM ORACLES SEPARATE THE POLYNOMIAL HIERARCHY

2.1. 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 [Furst et al. 1981]. 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 $n^{\Omega_d(\log n)}$ [Ajtai 1983]; 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 [Yao 1985], and shortly after Håstad proved the optimal lower bound of $\exp(\Omega(n^{1/(d-1)}))$ via his influential Switching Lemma [Håstad 1986a].

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 [Cai 1986]. Leveraging the connection made by [Furst et al. 1981], 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 [Babai 1987] gave a simpler proof of the same relativized separation.)

2.2. 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 [Baker et al. 1975]:

QUESTION 1. *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 [Furst et al. 1981]. It was first considered by Baker, Gill, and Solovay in their paper introducing the notion of relativiza-

tion [Baker et al. 1975], 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 [Baker and Selman 1979]. Following [Furst et al. 1981], Sipser noted the analogous connection between Meyer's question and circuit lower bounds [Sipser 1983]: 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 4 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 1.4), Håstad gave the first proof of a near-optimal n versus $\exp(n^{\Omega(1/d)})$ separation for the Sipser functions [Håstad 1986a], obtaining a strong depth hierarchy theorem for small-depth circuits and answering Meyer's question in the affirmative for all $d \in \mathbb{N}$.

2.3. 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 [Håstad 1986a; Håstad 1986b; Håstad 1989], Håstad poses the problem of extending his result to show this as an open question:

QUESTION 2 ([HÅSTAD 1986A; HÅSTAD 1986B; HÅSTAD 1989]). *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 2 also appears as the main open problem in [Cai 1986; Babai 1987]; as mentioned above, an affirmative answer to Question 2 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 2 comes from a surprising result of Book, who proved that the *unrelativized* polynomial hierarchy collapses if it collapses relative to a random oracle [Book 1994]. Over the years Question 2 has been discussed in a wide range of surveys [Johnson 1986; Hemaspaandra 1994; Shmoys and Tardos 1995; Hemaspaandra et al. 1995; Vollmer and Wagner 1997; Aaronson], textbooks [Du and Ko 2000; Hemaspaandra and Ogihara 2002], and research papers [Håstad 1986b; Håstad 1989; Tardos 1989; Fortnow 1999; Aaronson 2010a].

Our work. As a corollary of our main result (Theorem 1.1) — an *average-case* depth hierarchy theorem for small-depth circuits — we answer Question 2 in the affirmative for all $d \in \mathbb{N}$:

THEOREM 2.1 (THEOREM 1.2 RESTATED). *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 [Bennett and Gill 1981]. 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} [Aaronson 2010b; Aaronson 2010a]; he conjectures in [Aaronson 2010a] that his techniques can

be extended to resolve the $d = 2$ case of Theorem 1.2. We observe that O'Donnell and Wimmer's techniques (Theorem 1.5 in our introduction) can be used to reprove the $d = 1$ case [O'Donnell and Wimmer 2007], though the authors of [O'Donnell and Wimmer 2007] 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	[Furst et al. 1981]	[Sipser 1983]
Hard function(s)	Parity	Sipser functions
Relative to *some* oracle A	[Yao 1985; Håstad 1986a]	[Yao 1985; Håstad 1986a]
Relative to *random* oracle A	[Cai 1986; Babai 1987]	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 [Håstad 1986b] 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 1.2 follows from Theorem 1.1; see also [Rossman et al. 2015a]).

3. 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 [Linial et al. 1993]. 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 [Boppana 1997]:

THEOREM 3.1 (BOPPANA, LINIAL–MANSOUR–NISAN). *Let $f : \{0,1\}^n \to \{0,1\}$ be 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 3.1. Benjamini, Kalai and Schramm [Benjamini et al. 1999] conjectured that there is an absolute constant K such that every monotone Boolean function f is approximated on 99% of inputs by a depth-d circuit of size at most $\exp((K \cdot \mathbf{Inf}(f))^{1/(d-1)})$ for some $d \geq 2$. This conjecture was disproved by O'Donnell and Wimmer [O'Donnell and Wimmer 2007] (using the function from Theorem 1.5 in our introduction).

Following O'Donnell and Wimmer's disproof of the BKS conjecture, several researchers have posed questions similar in spirit. For example, O'Donnell asked if a weaker version of the BKS conjecture might be true with a relaxed bound on the size of the approximating circuit [O'Donnell 2007].

Our work. As a corollary of our main result, we provide a strong counterexample to versions of the BKS Conjecture in the bounded-depth and polylog influence regime (answering a question raised in Problem 4.6.3 of [Hatami 2014]).

THEOREM 3.2 (THEOREM 1.3 RESTATED).

(1) *For every constant d, there is a sequence of monotone functions $f^{(d)} : \{0,1\}^n \to \{0,1\}$ with total influence $\leq \log n$ such that $f^{(d)}$ cannot be approximated on 51% of inputs by depth-d circuits of size* $\exp(\exp(\Omega((\log n)^{1/d})))$.
(2) *There is a sequence of monotone functions $f : \{0,1\}^n \to \{0,1\}$ with total influence $\leq \log n$ such that f cannot be approximated on 51% of inputs by depth-$\sqrt{\log\log n}$ circuits of size* $\exp(\exp(\Omega(2^{\sqrt{\log\log n}})))$.

Theorem 1.3 shows that there is no *approximate converse* to Boppana–Linial–Mansour–Nisan (Theorem 3.1), which implies that depth-d circuits with total influence $\Omega(\log n)$ have size $\exp(\Omega(\log n)^{1/(d-1)})$ (i.e. exponentially weaker than bound (1)).

PROOF. For (1): Consider any fixed value of d. Let $f^{(d)} : \{0,1\}^n \to \{0,1\}$ be the depth $d+1$ Sipser function on the first $m = \exp(Cd(\log n)^{1/d})$ variables for a small constant $C > 0$ (to be determined). Since $f^{(d)}$ is computed by depth $d+1$ formulas of size m, a result of [Rossman 2015] (strengthening Theorem 3.1 for *formulas*) shows that

$$\mathbf{Inf}(f^{(d)}) = (O(\tfrac{1}{d}\log m))^d.$$

For an appropriate choice of constant C, we get $\mathbf{Inf}(f^{(d)}) \leq \log n$. On the other hand, by Theorem 1.1, depth-d circuits that agree with $f^{(d)}$ on 51% of inputs have size $\exp(m^{\Omega(1/d)}) = \exp(\exp(\Omega((\log n)^{1/d})))$.

For (2): With $d = \sqrt{\log\log n}$ and $m = \exp(Cd(\log n)^{1/d})$, we have $d = (\log m)^{o(1)}$. Therefore, Theorem 1.1 applies in this setting as well and we get the bound $\exp(\exp(\Omega((\log n)^{1/d}))) = \exp(\exp(\Omega(2^{\sqrt{\log\log n}})))$. □

4. OUR TECHNIQUES

The method of random restrictions dates back to Subbotovskaya [Subbotovskaya 1961] 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 [Furst et al. 1981; Ajtai 1983; Sipser 1983; Yao 1985; Håstad 1986a; Cai 1986; Babai 1987; Impagliazzo et al. 2012; Håstad 2014].

We begin in Section 4.1 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 4.2 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.

4.1. 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 $2 \leq k \leq d$, 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 Håstad's thesis [Håstad 1986b] 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.

4.2. 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 Section 7 for precise definitions of the restrictions that we will use.) 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 [Impagliazzo and Segerlind 2001].)

At a high level, our overall approach is structured around a sequence $\mathbf{\Psi}$ of carefully designed 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 6. Below 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 4.1):

– **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: 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 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 $\mathbf{\Psi}$. (This is in sharp contrast with our results, described below for Property 1, showing that the approximator "collapses to a simple function" with high probability under $\mathbf{\Psi}$.)
– **Property 1: Approximator C simplifies.** Finally, we prove that approximating circuits C of the types specified in our main lower bounds (Theorems 1.6 and 1.7) "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. Intuitively, the depth reduction of C follows by applying this lemma to every one of its bottom-level depth-2 subcircuits.

To put things slightly differently, our approach can be viewed in a stage-by-stage fashion. We show that after the i-th stage of random projections, the depth-d Sipser_d function is reduced to something similar to (but not exactly equal to) the depth-$(d-i)$ Sipser_{d-i} function, while the "approximating" circuit C has, with high probability, lost i levels.

4.3. Outline of the rest of the paper

We give basic definitions and set some terminology and notations in Section 5. In Section 6 we define the function Sipser_d. We define the precise restrictions and projections that we use in Section 7. Their key properties — that they generate uniformly random inputs and that they (with high probability and to a significant extent) transform Sipser_d into Sipser_{d-i} — are established in Section 8.

It is convenient for us to consider "i-th level random projections" for $i = 1, 2, \ldots, d-1$, and to view the i-th level random projection (denoted $\boldsymbol{\rho}^i$) as acting on gates at distance i from the input variables. A useful way to think about the draw of an i-th level random projection $\boldsymbol{\rho}^i$ is that first independent restrictions $\boldsymbol{\rho}^{i-1}$ are drawn for each sub-formula of depth $i-1$ (each gate at distance $i-1$ from the input variables) and then some additional fixing is done. This is also the way we reason about $\boldsymbol{\rho}^i$ when performing the simplifications, via a switching lemma, of the approximating circuit C. The switching lemma we require is stated and proved in Section 9. Its analysis is carried out using conditional probabilities using the formalism of [Håstad 1986b].

Finally, we put all the pieces together proving our main theorems in Section 10.

5. PRELIMINARIES

5.1. Notation and terminology

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.

A *restriction* ρ of a base set $\mathcal{X} = \{x_1, \ldots, x_n\}$ of Boolean variables is a function $\rho : [n] \to \{0, 1, *\}$ (we sometimes equivalently view a restriction ρ as a string $\rho \in \{0, 1, *\}^n$). Given a function $f(x_1, \ldots, x_n)$ and a restriction ρ, we write $f \restriction \rho$ to denote the function obtained by fixing x_i to $\rho(i)$ if $\rho(i) \in \{0, 1\}$ and leaving x_i unset if $\rho(i) = *$. For two restrictions $\rho, \tau \in \{0, 1, *\}^n$, we say that τ is a *refinement* of ρ if $\rho^{-1}(1) \subseteq \tau^{-1}(1)$ and $\rho^{-1}(0) \subseteq \tau^{-1}(0)$. In other words, every variable x_i 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). For two restrictions $\rho, \rho' \in \{0, 1, *\}^n$, their *composition*, denoted $\rho\rho' \in \{0, 1, *\}^n$, is the restriction defined by

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

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

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 $\{1, \ldots, k\}$. For F a Boolean function we write $\mathsf{depth}(F)$ to denote the minimal depth of any decision tree computing F

6. THE SIPSER FUNCTIONS

In this subsection we define the depth-d monotone n-variable read-once Boolean formula Sipser_d for $2 \leq d \in \mathbb{N}$ and establish some of its basic properties. The Sipser_d function is very similar to the depth-d formula considered by Håstad [Håstad 1986b]; the only difference is that the fan-ins of the gates at each level 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. For future reference it will be helpful to keep in mind that

$$m \approx \frac{\log n}{2d}.$$

Every leaf of Sipser_d occurs at the same depth (distance from the root) d; there are exactly n leaves (n will be defined in terms of m and d below) and each variable occurs at precisely one leaf. The formula is alternating; 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 $f_1, \ldots, f_{d-1}$, where f_i is the fan-in of every gate at distance i from the input variables. (We henceforth refer to such a gate as being at *level* i.)

The intuition that underlies the choice of fan-in sequence is that each gate should have the "right bias" when the n input variables are set according to a draw from the uniform distribution $\mathcal{U}$ over $\{0, 1\}^n$. More precisely, under such a draw we would like each internal (non-output) AND gate to be 1 with probability 2^{-2m} and each internal OR gate to be 1 with probability $1 - 2^{-2m}$, and we would like the output gate to be 1

with probabilty $1/2$. This cannot be achieved exactly but by an inductive choice of the fan-in sequence we can come very close. Let us turn to the formal details and give the parameters.

Definition 6.1. For $d \geq 2$, let $c_0 = \frac{1}{2}$ and for $1 \leq i \leq d-1$ let f_i be the smallest integer such that

$$(1 - c_{i-1})^{f_i} \leq 2^{-2m},$$

and set $c_i = (1 - c_{i-1})^{f_i}$. Finally set f_d to be the smallest integer such that

$$(1 - c_{d-1})^{f_d} \leq \frac{1}{2}.$$

As described above, the function Sipser_d is defined by an alternating depth-regular depth-d read-once monotone formula, with AND gates adjacent to the inputs, in which the fan-in of each gate at level i is f_i.

It is not difficult to see that

$$c_i \in [2^{-2m} - 2^{-4m}, 2^{-2m}] \text{ for } 1 \leq i \leq d-1 \tag{1}$$

while

$$\begin{aligned} f_1 &= 2m, \\ f_i &= 2m \ln 2 \cdot 2^{2m}(1 + O(2^{-m})) \text{ for } 2 \leq i \leq d-1, \\ f_d &= \ln 2 \cdot 2^{2m}(1 + O(2^{-m})). \end{aligned} \tag{2}$$

It follows that the number of inputs of F_d is

$$\prod_{i=1}^{d} f_i = 2^{2(d-1)m} m^{d-1} 2^{O(d)} \tag{3}$$

and we denote this number by n. We note that if $d \leq \frac{\log n}{2 \log \log n}$ then the first factor of (3) is the dominating factor and $m = \frac{\log n}{2d-2}(1 + o(1))$.

It follows by construction that if we feed a draw from $\mathcal{U}$ into the formula defining Sipser_d, then any gate at level i for odd i is an AND gate that is one with probability c_i, while any gate at level i for even i is an OR gate that is zero with probability c_i. It follows that the output of Sipser_d satisfies $|\mathbf{Pr}_{\boldsymbol{x} \sim \mathcal{U}}[\mathsf{Sipser}(x) = 1] - \frac{1}{2}| \leq 2^{-2m}$.

7. THE SPACES OF RANDOM PROJECTIONS $\mathcal{R}^1, \mathcal{R}^2, \ldots$

Our approach makes crucial use of projections, which are an extension of restrictions. Recall that a restriction maps each variable x_i to one of the three values 0, 1 and $*$. The two first values indicate that the corresponding constant should be substituted for this variables while the third value says that that the value of x_i remains undetermined.

Projections generalize restrictions in that a group of several variables may all be mapped to the same new variable. This makes further simplifications possible. In principle the mapping of old variables to new variables could be completely arbitrary, but to avoid a generality that we do not utilize we define only a special class of projections that are used in our proof.

The range set of variables that are mapped to by our projections is $\{x_v\}$ where v ranges over all nodes in the tree (formula) defining Sipser_d. Let V_i denote the set of variables x_v where v is at height i, i.e. at distance i from the inputs. Note that the set of original input variables to Sipser_d (i.e. the domain of our projections) is exactly given by V_0.

Definition 7.1. A level i projection, denoted ρ^i, is a mapping of V_0 into the set $\{0,1\} \cup V_i$. For $w \in [n]$ the possible values of $\rho^i(x_w)$ are 0, 1 and x_v where v is the height i ancestor of w in the Sipser_d formula.

The only way we construct a level i projection in this paper is to first do a level $i-1$ projection, then apply a classical restriction to the variables in V_{i-1}, and finally identify each remaining live variable with its parent. (Looking ahead, the formal definition of our distribution $\mathcal{R}^i$ of level i projections that we give below will be described using such a process.) Thus when going from a level $(i-1)$ projection ρ^{i-1} to a level i projection ρ^i we define a mapping ρ' from V_{i-1} to $\{0,1\} \cup V_i$ and ρ^i is the composition of ρ^{i-1} (a mapping from V_0 to $\{0,1\} \cup V_{i-1}$) with this ρ'. Any input mapped to a constant under ρ^{i-1} is still mapped to the same constant under ρ^i.

Our proof hinges on a sequence of carefully designed probability distributions $\mathcal{R}^1, \mathcal{R}^2, \ldots, \mathcal{R}^{d-1}$ where $\mathcal{R}^i$ is a distribution over level i projections; before giving formal definitions, let us explain the key properties of these random projections. A restriction $\boldsymbol{\rho}^i$ drawn from $\mathcal{R}^i$ operates independently on each height i sub-formula of Sipser_d; for the explanation below, let us assume that i is odd and hence that the top gate of such a sub-formula is an AND gate. (If i is even and the top gate is an OR gate we reverse the roles of 0 and 1 below.)

For $1 \leq i < d$, fix v to be any gate at level i in the Sipser_d formula. Let Sipser_v denote the sub-formula of Sipser_d that is rooted at v, and let $\mathsf{Inputs}(v) \subset V_0$ denote the subset of the original input variables to Sipser_d that belong to Sipser_v. The following (to be established later) are four key properties of our distribution $\mathcal{R}^i$ of level i projections.

KEY PROPERTIES: For $\boldsymbol{\rho}^i \leftarrow \mathcal{R}^i$,

(A) With probability $2^{-5m/2}$, all variables of $\mathsf{Inputs}(v)$ are fixed to constants in such a way that $\mathsf{Sipser}_v \restriction \boldsymbol{\rho}^i \equiv 1$.
(B) With probability $1 - 2^{-m}$, all variables of $\mathsf{Inputs}(v)$ are fixed to constants in such a way that $\mathsf{Sipser}_v \restriction \boldsymbol{\rho}^i \equiv 0$.
(C) With the remaining probability $2^{-m} - 2^{-5m/2}$, we have $\mathsf{Sipser}_v \restriction \boldsymbol{\rho}^i \equiv x_v$.
(D) If x_v is set to 1 with probability b_i defined as

$$b_i = \frac{c_i - 2^{-5m/2}}{2^{-m} - 2^{-5m/2}} \tag{4}$$

and is set to 0 with the remaining probability $1 - b_i$, then $\boldsymbol{\rho}^i$ combined with this setting gives a uniformly random assignment to all variables in $\mathsf{Inputs}(v)$.

We remark that properties (A)–(C) correspond to $\mathsf{Sipser} \restriction \boldsymbol{\rho}^i$ "retaining structure" as discussed in Sections 4 and 4.2, while property (D) corresponds to our random projections "completing to the uniform distribution."

For future reference we note that

$$b_i = 2^{-m}(1 - O(2^{-m/2})). \tag{5}$$

Having presented the key properties of the $\mathcal{R}^i$'s, we now proceed to a formal description of these distributions. We begin by defining $\mathcal{R}^1$ and then continue to the definition of a general $\mathcal{R}^i$ (which is defined recursively using $\mathcal{R}^{i-1}$).

7.1. Definition of the space of random projections $\mathcal{R}^1$

Recall that for v a gate at level one of Sipser_d, the set $\mathsf{Inputs}(v)$ consists of those original input variables that lie below v (so $\mathsf{Inputs}(v)$ is a set of size $2m$). Recall also that a level one projection (i.e. an element of the support of $\mathcal{R}^1$) is a mapping from V_0 to $\{0,1\} \cup V_1$ such that each original input variable x_w is mapped either to $\{0,1\}$ or to x_v where v is the parent of w.

Definition 7.2. A random restriction $\boldsymbol{\rho}^1 \leftarrow \mathcal{R}^1$ is generated by doing the following independently for each gate v at level one. First draw a random restriction $\boldsymbol{\rho} \in \{0,1,*\}^{\mathsf{Inputs}(v)}$ as follows:

(1) Draw a uniform random assignment $\boldsymbol{\alpha} \in \{0,1\}^{2m}$ to all inputs $x_w \in \mathsf{Inputs}(v)$.
(2) If $\boldsymbol{\alpha} = 1^{2m}$ then with probability $2^{-m/2}$ set $\boldsymbol{\rho}(x_w) = 1$ for all $x_w \in \mathsf{Inputs}(v)$, and otherwise proceed as follows. Pick a uniformly random subset $\mathbf{S}$ of $\mathsf{Inputs}(v)$ conditioned on $\mathbf{S} \neq \emptyset$ and set $\boldsymbol{\rho}(x_w) = *$ for $x_w \in \mathbf{S}$ and $\boldsymbol{\rho}(x_w) = 1$ for $x_w \notin \mathbf{S}$.
(3) Otherwise (if $\boldsymbol{\alpha} \neq 1^{2m}$), with probability $(1-2^{-m})/(1-2^{-2m})$ set $\boldsymbol{\rho}(x_w) = \boldsymbol{\alpha}_w$ for all $x_w \in \mathsf{Inputs}(v)$ and otherwise set $\boldsymbol{\rho}(x_w) = *$ for all w such that $\boldsymbol{\alpha}_w = 0$ while $\boldsymbol{\rho}(x_w) = 1$ for all w such that $\boldsymbol{\alpha}_w = 1$.

The random projection $\boldsymbol{\rho}^1$ is obtained by identifying all variables $x_w \in \mathsf{Inputs}(v)$ such that $\boldsymbol{\rho}(x_w) = *$ with x_v (i.e. $\boldsymbol{\rho}^1$ maps all such x_w to x_v).

The following may be helpful intuition to aid in understanding Definition 7.2. We think of the assignment $\boldsymbol{\alpha}$ as a "tentative" assignment to all variables. Steps (2) and (3) "forget" some of the values, but as we show below, if we later assign x_v an element of $\{0,1\}$ with the correct bias then we recover the same probability distribution as if we had kept the original $\boldsymbol{\alpha}$. This assures that such a substitution creates a uniformly random input.

Let ρ^1 be a restriction in the support of $\mathcal{R}^1$. For a level one gate v, if any variable $x_w \in \mathsf{Inputs}(\mathsf{Sipser}_d)$ is mapped to x_v by ρ then we say that x_v is *alive* and also write this as $\rho^1(x_v) = x_v$. Keeping with this convention we also write $\rho^1(x_v) = c$ when Sipser_v is fixed to the constant $c \in \{0,1\}$ by ρ^1. In general we sometimes write $\rho^1(x_v)$ for $\mathsf{Sipser}_v \restriction \rho^1$ and remember that this takes values 0, 1 or x_v.

In the rest of this subsection we verify that $\mathcal{R}^1$ indeed has the four key properties (A)–(D) described earlier. These verifications are mostly straightforward but let us still check the properties in detail.

For (A), we observe that the probability that Sipser_v is fixed to 1 under $\boldsymbol{\rho}^1$ is $2^{-5m/2}$, as we need to pick $\boldsymbol{\alpha} = 1^{2m}$ and then decide to use this assignment fully in step 2. Similarly, for (B) the probability that Sipser_v is fixed to 0 under $\boldsymbol{\rho}^1$ is

$$(1-2^{-2m}) \cdot (1-2^{-m})/(1-2^{-2m}) = 1-2^{-m}$$

as this must happen in step 3. Note that in all other cases we have a non-empty set $\mathbf{S}$ such that $\boldsymbol{\rho}(x_w) = x_v$ for all $x_w \in \mathbf{S}$ while all other variables of $\mathsf{Inputs}(v)$ are mapped to 1. This implies that $\mathsf{Sipser}_v \restriction \rho^1 = x_v$, giving (C).

It remains to establish (D), which follows from the following lemma:

LEMMA 7.3. *Let $\boldsymbol{\rho}^1 \leftarrow \mathcal{R}^1$. Let $\boldsymbol{\rho}'$ be the refinement of $\boldsymbol{\rho}^1$ obtained as follows: for each x_v that is alive under $\boldsymbol{\rho}^1$, set it to 1 with probability $b_1 = \frac{2^{-2m}-2^{-5m/2}}{2^{-m}-2^{-5m/2}}$ and to 0 with probability $1-b_1$. Then we have that $\boldsymbol{\rho}'$ is distributed as a uniform random 0/1 assignment to $\mathsf{Inputs}(v)$.*

PROOF. Fix any gate v at level one. For any non-empty subset $S \subseteq \mathsf{Inputs}(v)$, the probability that S is chosen to receive $*$'s in step 2 is

$$p_2^S = \frac{2^{-2m}(1-2^{-m/2})}{2^{2m}-1}$$

while the probability that the same set S is chosen in step 3 is

$$p_3^S = \frac{2^{-2m}(2^{-m}-2^{-2m})}{(1-2^{-2m})} = \frac{(2^{-m}-2^{-2m})}{(2^{2m}-1)}.$$

This implies that, conditioned on S being the set of input variables x_w such that $\boldsymbol{\rho}^1(x_w) = x_v$, the probability that S was chosen in step 2 is

$$\frac{p_2^S}{p_2^S + p_3^S} = \frac{2^{-2m} - 2^{-5m/2}}{2^{-m} - 2^{-5m/2}} \tag{6}$$

and this is exactly b_1. This implies that if x_v is set to 1 with probability b_1 (and to 0 with probability $1 - b_1$), then we get the same probability distribution on assignments to $\mathsf{Inputs}(v)$ as if we had set $x_w = \boldsymbol{\alpha}_w$ immediately. We conclude that we get a uniformly random distribution over all 2^{2m} assignments to $\mathsf{Inputs}(v)$. □

Having shown that R^1 has all four of the key properties, we proceed to the case of a general $\mathcal{R}^i$.

7.2. Definition of the space of random projections $\mathcal{R}^i$ for general $i > 1$

When drawing a projection from the distribution $\mathcal{R}^i$ we first draw a projection $\boldsymbol{\rho}^{i-1} \leftarrow \mathcal{R}^{i-1}$ which reduces each sub-formula Sipser_w of depth $i - 1$ to either a constant or a variable x_w. As a result, going from $\boldsymbol{\rho}^{i-1} \leftarrow \mathcal{R}^{i-1}$ to $\boldsymbol{\rho}^i \leftarrow \mathcal{R}^i$ is quite similar to drawing a projection from $\mathcal{R}^1$. We again center the construction around a Boolean vector $\boldsymbol{\alpha}$ which now plays the role of a vector of independent but suitably biased values at level $i - 1$. The bits of $\boldsymbol{\alpha}$ now come in two flavors: there are "hard" bits which should be thought of as already fixed by $\boldsymbol{\rho}^{i-1}$ and hence cannot be changed to x_v, and the other bits are "soft" bits that can be changed.

As noted earlier, the random projection $\boldsymbol{\rho}^i \leftarrow \mathcal{R}^i$ is independent across all nodes v at level i, so to describe a draw of $\boldsymbol{\rho} \leftarrow \mathcal{R}^i$ it suffices to fix a level i gate v and explain how $\boldsymbol{\rho}^i$ is generated at v. In the following explanation we assume that i is odd and hence that each gate v on level i is an AND gate (in the case of even i each gate v on level i is an OR gate and the roles of 0 and 1 are reversed in what follows). For v a gate at level i we let $\mathsf{Children}(v)$ be the set of gates that feed directly into v, so $\mathsf{Children}(v)$ is of size f_i. (Note that $\mathsf{Children}(v)$ plays the role that $\mathsf{Inputs}(v)$ played when v was a level one gate; indeed, for v a level one gate $\mathsf{Children}(v)$ and $\mathsf{Inputs}(v)$ are the same set, though they are different sets at levels greater than one.)

We first define a distribution $\mathcal{D}_i$ that is used to pick an input $\boldsymbol{\alpha} \in \{0,1\}^{f_i}$ where some values are hard while other are soft. A draw of $\boldsymbol{\alpha} \leftarrow \mathcal{D}_i$ is obtained as follows: for each coordinate w of $\boldsymbol{\alpha}$ independently (i.e. for each $w \in \mathsf{Children}(v)$),

(1) Make $\boldsymbol{\alpha}_w$ a *hard zero* with probability $2^{-5m/2}$.
(2) Make $\boldsymbol{\alpha}_w$ a *hard one* with probability $1 - 2^{-m}$.
(3) Make $\boldsymbol{\alpha}_w$ a *soft zero* with probability $c_{i-1} - 2^{-5m/2}$.
(4) Make $\boldsymbol{\alpha}_w$ a *soft one* with probability $2^{-m} - c_{i-1}$.

We note that each coordinate that is not given a hard value is set to a soft zero with probability exactly b_{i-1}.

Given a draw of $\boldsymbol{\alpha} \leftarrow \mathcal{D}_i$, we write $\mathbf{Soft}(\boldsymbol{\alpha})$ to denote the subset of $\mathsf{Children}(v)$ which is the set of coordinates that are given soft values. Thus typically $\mathbf{Soft}(\boldsymbol{\alpha})$ is of size roughly $2^{-m} f_i$; we write $\boldsymbol{f}_v(\boldsymbol{\alpha}) = |\mathbf{Soft}(\boldsymbol{\alpha})|$ to denote the actual size of $\mathbf{Soft}(\boldsymbol{\alpha})$.

Given a subset $T \subseteq \mathsf{Children}(v)$ (which should be thought of as a possible outcome of $\mathbf{Soft}(\boldsymbol{\alpha})$), let $\mathcal{P}_T$ denote the following probability distribution over nonempty subsets of T: a draw of $\mathbf{S} \leftarrow \mathcal{P}_T$ (which should be thought of as a potential set of soft zeros) is obtained by independently including each element of T with probability b_{i-1}, and if $\mathbf{S}$ comes out to be empty we try again. Note that if T is about the typical size $2^{-m} f_i$ of $\mathbf{Soft}(\boldsymbol{\alpha})$, then the probability that $\mathbf{S}$ is empty is about 2^{-2m} so this conditioning is in general mild. (We will sometimes refer to a $\mathbf{S} \leftarrow \mathcal{P}_T$ as a "uniform non-empty subset of T of bias b_{i-1}.") For fixed sets $\emptyset \neq S \subseteq T$ let us write $q_{S,T}$ as shorthand for

$\mathbf{Pr}_{\mathbf{S}\leftarrow\mathcal{P}_T}[\mathbf{S}=S]$. Then it is not difficult to see that

$$q_{S,T} = (1-(1-b_{i-1})^{|T|})^{-1} b_{i-1}^{|S|}(1-b_{i-1})^{|T|-|S|}. \tag{7}$$

With this setup out of the way, we are ready to describe how $\boldsymbol{\rho}^i \leftarrow \mathcal{R}^i$ is generated at v. For notational convenience in our later arguments we first describe how a closely related restriction $(\boldsymbol{\rho}')^i$ is generated at v (from a distribution over $\{0,1,*\}^{\mathsf{Children}(v)}$); in the last step of the draw, $\boldsymbol{\rho}^i$ (which belongs to $\{0,1,x_v\}^{\mathsf{Children}(v)}$) is obtained from $(\boldsymbol{\rho}')^i$ by a simple variable identification (projection).

We proceed to the definition of $\mathcal{R}^i$. (The following is analogous to and should be compared with Definition 7.2.) We remind the reader again that i is assumed to be odd and hence v is an AND gate. A draw of $(\boldsymbol{\rho}')^i$ and $\boldsymbol{\rho}^i$ from $\mathcal{R}^i$ is obtained as follows: [1]

$(0)^i$ First draw $\boldsymbol{\alpha} \leftarrow \mathcal{D}_i$ as described above.
(Looking ahead, we never change a hard value while soft values are either made permanent or may be turned into x_v. We further note that the cases $(1)^i$ – $(4)^i$ below should be viewed as disjoint and mutually exclusive, i.e. the description of each case implicitly assumes that none of the previous cases hold.)

$(1)^i$ If $\boldsymbol{\alpha}$ has at least one hard zero coordinate, then set $(\boldsymbol{\rho}')^i(x_w) = \boldsymbol{\alpha}_w$ for all $w \in \mathsf{Children}(v)$.
(Intuitively, if there is a hard zero present then the value of Sipser_v is fixed to zero.)

$(2)^i$ If $|\boldsymbol{f}_v(\boldsymbol{\alpha}) - f_i 2^{-m}| \geq 2^{3m/4}$ then set $(\boldsymbol{\rho}')^i(x_w) = \alpha_w$ for all w.
(Intuitively, if a very atypical number of soft values are received, then we give up on target preservation. This is okay because as we will see it is a very low probability event to have such an atypical number of soft values.)

Let us turn to the more interesting cases in the definition of $(\boldsymbol{\rho}')^i$. Let q_3 and $q_4(\boldsymbol{f}_v(\boldsymbol{\alpha}))$ be constants, which in rough terms satisfy $q_3 \approx 2^{-m/2}$ and $q_4(\boldsymbol{f}_v(\boldsymbol{\alpha})) \approx 2^{-m}$, but whose exact values are given during the analysis in Section 8.

$(3)^i$ If $\boldsymbol{\alpha} = 1^{f_i}$, then with probability q_3 set $(\boldsymbol{\rho}')^i(x_w) = 1$ for all $w \in \mathsf{Children}(v)$ and with the remaining $1-q_3$ probability proceed as follows. Choose a non-empty subset $\mathbf{S}$ of $\mathbf{Soft}(\boldsymbol{\alpha})$ with bias b_{i-1} (i.e. draw $\mathbf{S} \leftarrow \mathcal{P}_{\mathbf{Soft}(\boldsymbol{\alpha})}$) and set $(\boldsymbol{\rho}')^i(x_w) = *$ for $x_w \in \mathbf{S}$ and $(\boldsymbol{\rho}')^i(x_w) = 1$ for $x_w \notin \mathbf{S}$.
(Note the similarity to Step 2 of Definition 7.2.)

$(4)^i$ If $\mathbf{Soft}(\boldsymbol{\alpha})$ is nonempty then define $\mathbf{S}$ to equal $\mathbf{Soft}(\boldsymbol{\alpha})$. With probability $1-q_4(\boldsymbol{f}_v(\boldsymbol{\alpha}))$ set $(\boldsymbol{\rho}')^i(x_w) = \boldsymbol{\alpha}_w$ for all $w \in \mathsf{Children}(v)$ and with the remaining $q_4(\boldsymbol{f}_v(\boldsymbol{\alpha}))$ probability set $(\boldsymbol{\rho}')^i(x_w) = *$ for $w \in \mathbf{S}$ and set $(\boldsymbol{\rho}')^i(x_w) = 1 = \alpha_w$ for $w \notin \mathbf{S}$.
(Note the similarity to Step 2 of Definition 7.2.)

[2]

This concludes the description of how $(\boldsymbol{\rho}')^i$ is drawn from $\mathcal{R}^i$. Finally, the random projection $\boldsymbol{\rho}^i$ is obtained from $(\boldsymbol{\rho}')^i$ by identifying all of the live variables $x_w \in \mathsf{Children}(v)$ (i.e. all of the x_w such that $(\boldsymbol{\rho}')^i(x_w) = *$) with x_v (i.e. $\boldsymbol{\rho}^i$ maps all such x_w to x_v).

Before proceeding let us observe that if any coordinate $\boldsymbol{\alpha}_w$ is set to a hard value, then x_w is always set to this value under $\boldsymbol{\rho}^i$. This follows as in both steps $(3)^i$ and $(4)^i$, $\mathbf{S}$ is a subset of $\mathbf{Soft}(\boldsymbol{\alpha})$, and hence is not given hard ones as values, and if any hard zero is assigned then all values of $\boldsymbol{\alpha}$ are used in $\boldsymbol{\rho}^i$.

[1] **Rocco:** The format below is that intuition is mixed in with the algorithmic description of the draw from the distribution, with the intuition in parentheses on a new line below each algorithmic step. This has the advantage of having the intuition for each step be right by the description of the step but has the drawback of mixing the intuition up with the formal description.

[2] **Rocco:** Check carefully: this $\boldsymbol{\rho}^i$ just defined maps what variables to what? Is it actually $\mathcal{R}^i$ or should it be called something else? It hasn't been coupled with $\mathcal{R}^{i-1}$ yet...

Steps $(1)^i$ – $(4)^i$ above tell us how to go from $\boldsymbol{\alpha}$ to $\boldsymbol{\rho}^i$. As indicated at the start of this subsection, the hard versus soft status of the bits of $\boldsymbol{\alpha}$ is what ties $\mathcal{R}^i$ to the earlier distributions $\mathcal{R}^{i-1}$ since as mentioned above hard bits of $\boldsymbol{\alpha}$ should be thought of as already fixed by some $\boldsymbol{\rho}^{i-1}$ drawn from $\mathcal{R}^{i-1}$. There are a couple of equivalent ways to view the combination of draws $\boldsymbol{\rho}^{i-1} \leftarrow \mathcal{R}^{i-1}$ and $\boldsymbol{\rho}^i \leftarrow \mathcal{R}^i$ as specifying a full assignment (i.e. to complete the overall recursive definition of $\mathcal{R}^i$). One way is the following:

(1) Draw an assignment $\boldsymbol{\alpha}$ from $\mathcal{D}_i$.
(2) For each hard coordinate w of $\boldsymbol{\alpha}$, independently draw a restriction $\boldsymbol{\rho}^{i-1} \leftarrow \mathcal{R}^{i-1}$ conditioned on $\boldsymbol{\rho}^{i-1}(x_w)$ equalling this bit value $\boldsymbol{\alpha}_w$.
(3) For each soft coordinate w of $\boldsymbol{\alpha}$, independently draw a restriction $\boldsymbol{\rho}^{i-1} \leftarrow \mathcal{R}^{i-1}$ conditioned on $\boldsymbol{\rho}^{i-1}(x_w) = x_w$, and then set $\boldsymbol{\rho}^i(x_w)$ as described in steps $(1)^i$ – $(4)^i$ above. [3]

An equivalent way to describe the procedure is as follows: First draw a random independent $\boldsymbol{\rho}^{i-1} \leftarrow \mathcal{R}^{i-1}$ for each depth $i-1$ gate $w \in \mathsf{Children}(v)$, and then draw a value of $\boldsymbol{\alpha}$ conditioned on getting hard coordinates with the correct value for each $\boldsymbol{\rho}^{i-1}(x_w)$ that was chosen to be a constant. In more detail this is the following procedure:

(I) Draw a random $\boldsymbol{\rho}^{i-1} \in \mathcal{R}^{i-1}$. For each level $i-1$ gate w such that $\boldsymbol{\rho}^{i-1}(x_w)$ is a constant, fix $\boldsymbol{\alpha}_w$ to be that constant in a hard way.
(II) For each level $i-1$ gate w such that $\boldsymbol{\alpha}_w$ is not set in step 1, pick it to be a soft zero with probability b_{i-1} and a soft one with the remaining $1-b_{i-1}$ probability.

At this point an assignment from $\{0,1\}^{\mathsf{Children}(v)}$ has been obtained for $\boldsymbol{\alpha}$. Based on this assignment, continue as follows:

(III) For any $w \in \mathsf{Children}(v)$ such that case $(1)^i$ or $(2)^i$ applies, fix the value of x_w to a constant based on those cases. This is done by a traditional restriction taking values $0, 1$ and $*$ (i.e. a distribution over $\{0,1,*\}^{\mathsf{Children}(v)}$) and we denote this random restriction by $\boldsymbol{\rho}_1$ and denote its distribution by $\mathcal{R}_1^i$. (To be explicit, a variable x_w gets assigned $*$ under $\boldsymbol{\rho}_1 \leftarrow \mathcal{R}_1^i$ if neither case $(1)^i$ nor case $(2)^i$ applies to it).
(IV) Given the outcome of $\boldsymbol{\rho}_1$, we extend it to obtain a restriction $\boldsymbol{\rho}_2$ [4] as follows: for any $w \in \mathsf{Children}(v)$ such that case $(3)^i$ or $(4)^i$ applies, fix the value of x_w to a constant based on those cases. This is again done by a traditional restriction taking values $0, 1$ and $*$ and we denote this random restriction (which extends $\boldsymbol{\rho}_1$) by $\boldsymbol{\rho}_2$ and denote its distribution by $\mathcal{R}_2^i(\boldsymbol{\rho}_1)$. (Again to be explicit, a variable x_w gets assigned $*$ under $\boldsymbol{\rho}_2 \leftarrow \mathcal{R}_2^i$ if neither case $(3)^i$ nor case $(4)^i$ applies to it).
(V) Finally, $\boldsymbol{\rho}^i \leftarrow \mathcal{R}_i$ is defined as follows: for each $w \in \mathsf{Children}(v)$ that has $\boldsymbol{\rho}_2(x_w) = *$, we take $\boldsymbol{\rho}^i(x_w) = x_v$ (and for each $w \in \mathsf{Children}(v)$ that has $\boldsymbol{\rho}_2(x_w) \in \{0,1\}$ we take $\boldsymbol{\rho}^i(x_w) = \boldsymbol{\rho}_2(x_w)$). We let π denote this final projection step that maps each live x_w to x_v. (Note that π is not boldfaced as there is no randomness involved — it deterministically maps every x_w to x_v for every $w \in \mathsf{Children}(v)$.)

We say that $\boldsymbol{\rho}^{i-1}$, $\boldsymbol{\rho}_1$, $\boldsymbol{\rho}_2$ and π are the *components* of $\boldsymbol{\rho}^i$. The most interesting part when going from $\boldsymbol{\rho}^{i-1}$ to $\boldsymbol{\rho}^i$ turns out to be the third step $\boldsymbol{\rho}_2$. For a function f we let $f \restriction \boldsymbol{\rho}_2$ denote the function after this step. We let $f \restriction (\boldsymbol{\rho}_2 \circ \pi)$ denote the function after the projection π has also been applied. We observe that $f \restriction \boldsymbol{\rho}_2$ is a function of the variables in V_{i-1} while $f \restriction \boldsymbol{\rho}_2 \circ \pi$ is a function of the variables in V_i.

[3] **Rocco:** Is this right?
[4] **Rocco:** Note: renamed what was previously "ρ" as $\boldsymbol{\rho}_2$

As an example suppose that $f = x_{w_1} \vee \bar{x}_{w_2}$ where w_1 and w_2 are two nodes in the same level i sub-formula Sipser_v for some level i node v. Suppose furthermore that ρ_2 is an outcome of $\boldsymbol{\rho}_2$ that does not fix either of these variables. In this situation $f \restriction \rho_2$ is the same function as f while $f \restriction (\boldsymbol{\rho}_2 \circ \pi) = x_v \vee \bar{x}_v$ is identically true.

8. THE DISTRIBUTIONS $\mathcal{R}^I$ SATISFY THE KEY PROPERTIES

The construction of the distribution of restrictions $\mathcal{R}^i$ has been carefully crafted to, more or less by definition, satisfy the four key properties. In this section we show that these properties hold via a sequence of lemmas, starting with the following simple observation.

LEMMA 8.1. *Let ρ^i be a projection in the support of $\mathcal{R}^i$ and let v be a level i node in Sipser_d such that x_v is alive under $\boldsymbol{\rho}^i$. Then we have $\mathsf{Sipser}_v \restriction \rho^i = x_v$. Furthermore if $\mathsf{Sipser}_v \restriction \rho^i$ is a constant then ρ^i assigns constants to all variables in Sipser_v.*

PROOF. Going over the construction line by line it is not difficult to see that this is true. □

The second lemma says that each level i sub-formula is projected to 1 with the correct probability under $\mathcal{R}^i$:

LEMMA 8.2 (KEY PROPERTY (A)). *There is a value of $q_3 = 2^{-m/2}(1 + o(1))$ [5] such that $\mathbf{Pr}_{\boldsymbol{\rho}^i \leftarrow \mathcal{R}^i}[\mathsf{Sipser}_v \restriction \boldsymbol{\rho}^i = 1] = 2^{-5m/2}$.*

PROOF. Let p_2 denote the probability that that case $(2)^i$ happens when $\boldsymbol{\rho}^i$ is drawn from $\mathcal{R}^i$. By standard Chernoff bounds [6] we have $p_2 = \exp(-\Omega(2^{m/2}/m))$. Let $p_{2,1}$ be the probability that the value of Sipser_v is fixed to 1 under case $(2)^i$ (so $p_{2,1} \leq p_2$ is extremely small).

Now let p_3 denote the probability that case $(3)^i$ happens. [7] Since $\mathbf{Pr}_{\boldsymbol{\alpha} \leftarrow \mathcal{D}_i}[\boldsymbol{\alpha} = 1^{f_i}] = c_i$, we have that $p_3 = c_i - p_{2,1}$ and thus $p_3 = 2^{-2m}(1 - o(1))$. Fix the value of q_3 to be $(2^{-5m/2} - p_{2,1})/p_3$ and note that $q_3 = 2^{-m/2}(1 + o(1))$ as promised. The probability that Sipser_v is fixed to the value 1 under $\boldsymbol{\rho}^i \leftarrow \mathcal{R}^i$ is $p_{2,1} + p_3 q_3$ and this, by the choice of q_3, equals $2^{-5m/2}$. □

We next determine a suitable value for $q_4(\boldsymbol{f}_v(\boldsymbol{\alpha}))$. (Lemmas 8.3 and 8.4 can together be viewed as establishing that the random projections $\mathcal{R}^i$ "complete to the uniform distribution.")

LEMMA 8.3 (TOWARDS KEY PROPERTY (D)). *Let v be a level i node in Sipser_d. There is a value of $q_4(\boldsymbol{f}_v(\boldsymbol{\alpha})) = 2^{-m}(1 + o(1))$ such that setting x_v to 1 with probability b_i (and to 0 with probability $1 - b_i$) after we have drawn a random $\boldsymbol{\rho}^i \leftarrow \mathcal{R}^i$, gives the same distribution over assignments to the variables $\{x_w : w \in \mathsf{Children}(v)\}$ as setting $x_w = \boldsymbol{\alpha}_w$, $\boldsymbol{\alpha} \leftarrow \mathcal{D}_i$, for all $w \in \mathsf{Children}(v)$.*

PROOF. We prove that for any $\emptyset \neq S \subseteq T$, conditioned on T being the set $\mathbf{Soft}(\boldsymbol{\alpha})$ of soft values obtained when $\boldsymbol{\alpha} \leftarrow \mathcal{D}_i$ and S being the set $\mathbf{S}$ of variables set to x_v in case $(3)^i$ or case $(4)^i$, the probability that this happened in case $(3)^i$ is b_i while the probability that this happened in case $(4)^i$ is $1 - b_i$. Since we in case $(3)^i$ change the values of the variables in S from 1 to x_v and in case $(4)^i$ change the values from 0 to x_v, as in the proof of Lemma 7.3 this is sufficient to prove Lemma 8.3. For the rest of

[5] **Rocco:** Here and elsewhere where "$1 + o(m)$" appears, is this really the best thing to say – it seems a bit odd to specify a lower-order term and then say "$+o(m)$"
[6] **Rocco:** check
[7] **Rocco:** check math in next sentence or two; shouldn't p_3 be less than c_i which is less than 2^{-2m}?

this proof we condition on $\boldsymbol{\alpha}$ being such that $\mathbf{Soft}(\alpha) = T$ for a particular non-empty set T of size $\boldsymbol{f}_v(\boldsymbol{\alpha})$ with $|\boldsymbol{f}_v(\boldsymbol{\alpha}) - f_i 2^{-m}| < 2^{3m/4}$ and no hard zero being picked.

The conditional probability of having $\mathbf{S} = S$ and reaching case $(3)^i$ is

$$p^3_{S,T} = (1 - b_{i-1})^{\boldsymbol{f}_v(\boldsymbol{\alpha})} q_{S,T}(1 - q_3),$$

where $q_{S,T}$ is the probability as in (7). The probability of getting the same sets in case 4 is

$$p^4_{S,T} = (1 - (1 - b_{i-1})^{\boldsymbol{f}_v(\boldsymbol{\alpha})}) q_{S,T} q_4(\boldsymbol{f}_v(\boldsymbol{\alpha})).$$

We now set [8]

$$q_4(\boldsymbol{f}_v(\boldsymbol{\alpha})) = \frac{(1 - b_{i-1})^{\boldsymbol{f}_v(\boldsymbol{\alpha})}(1 - b_i)(1 - q_3)}{(1 - (1 - b_{i-1})^{\boldsymbol{f}_v(\boldsymbol{\alpha})}) b_i} \tag{8}$$

with the result that

$$\frac{p^4_{S,T}}{p^3_{S,T}} = \frac{1 - b_i}{b_i}. \tag{9}$$

By the given bounds on $\boldsymbol{f}_v(\boldsymbol{\alpha})$ we know that $\boldsymbol{f}_v(\boldsymbol{\alpha}) = f_i 2^{-m}(1 + O(2^{-m/4}))$ and hence $(1 - b_{i-1})^{\boldsymbol{f}_v(\boldsymbol{\alpha})} = 2^{-2m}(1 + o(1))$. Furthermore as $b_i = 2^{-m}(1 + o(1))$ it is possible to satisfy (8) with $q_4(\boldsymbol{f}_v(\boldsymbol{\alpha})) = 2^{-m}(1 + o(1))$. □

From now on we assume that we use the values of q_3 and $q_4(\boldsymbol{f}_v(\boldsymbol{\alpha}))$ determined by Lemma 8.2 and Lemma 8.3. The next lemma is, more or less, an immediate consequence of Lemma 8.3.

LEMMA 8.4 (KEY PROPERTY (D)). *Let v be a level i node in Sipser_d. Then drawing $\boldsymbol{\rho}^i \leftarrow \mathcal{R}^i$ and then setting x_v to 1 with probability b_i (and to 0 with probability $1 - b_i$) gives the uniform distribution on $\mathsf{Inputs}(v)$.*

PROOF. We proceed by induction on i, noting that for $i = 1$ we already established the desired result in Lemma 7.3.

Lemma 8.3 tells us that fixing x_v to 1 with probability b_i (and to 0 with the remaining probability) is the same as setting the soft values according to $\boldsymbol{\alpha}$. This, in its turn, is the same as picking independent restrictions from $\boldsymbol{\rho}^{i-1}$ for each sub-formula Sipser_w and setting any live x_w to 1 with probability $1 - b_{i-1}$ and to 0 with probability b_{i-1}. By induction this results in the uniform distribution. □

Finally, we verify key properties (B) and (C) of $\mathcal{R}^i$.

LEMMA 8.5 (KEY PROPERTIES (B) AND (C)). *Fix a level i node v. We have $\mathbf{Pr}_{\boldsymbol{\rho}^i \leftarrow \mathcal{R}^i}[\mathsf{Sipser}_v \restriction \boldsymbol{\rho}^i = 0] = 1 - 2^{-m}$ and $\mathbf{Pr}_{\boldsymbol{\rho}^i \leftarrow \mathcal{R}^i}[\mathsf{Sipser}_v \restriction \boldsymbol{\rho}^i \equiv x_v] = 2^{-m} - 2^{-5m/2}$.*

PROOF. This could be done by a tedious calculation, but in fact it can be seen by a high level argument. The restriction $\boldsymbol{\rho}^i$ can reduce Sipser_v to 0, 1 or x_v. Lemma 8.2 says that the value 1 is taken with the correct probability, and Lemma 8.4 says that if x_v is set to 1 with probability b_i then we get a uniformly random input and hence the output of Sipser_v is one with probability c_i. This implies that

$$2^{-5m/2} + b_i \, \mathbf{Pr}_{\boldsymbol{\rho}^i \leftarrow \mathcal{R}^i}[\mathsf{Sipser}_v \restriction \boldsymbol{\rho}^i \equiv x_v] = c_i$$

and hence, by the definition of b_i, we conclude that $\mathbf{Pr}_{\boldsymbol{\rho}^i \leftarrow \mathcal{R}^i}[\mathsf{Sipser}_v \restriction \boldsymbol{\rho}^i \equiv x_v] = 2^{-m} - 2^{-5m/2}$ as desired. Since the probabilities of obtaining the three possible values for $\mathsf{Sipser}_v \restriction \boldsymbol{\rho}^i$ sum to one, the lemma follows. □

[8] **Rocco:** check this math

The most interesting property of our restrictions is that we can prove a switching lemma and we proceed with this step.

9. THE SWITCHING LEMMAS

To establish a general hierarchy result (in particular one that separates depth d from depth $d-2$) it is sufficient to prove a switching lemma for $\mathcal{R}^i$ for $i \geq 2$, and in view of this we prove this lemma first. To get a tight result we later prove a modified lemma that applies to $\mathcal{R}^1$.

As discussed in Section 7, a restriction $\boldsymbol{\rho}^i \leftarrow \mathcal{R}^i$ is chosen by first drawing $\boldsymbol{\rho}^{i-1} \leftarrow \mathcal{R}^{i-1}$, followed by $\boldsymbol{\rho}_1$, $\boldsymbol{\rho}_2$ and finally making a projection π. In this section we assume any fixed values of of ρ^{i-1} for $\boldsymbol{\rho}^{i-1}$ and ρ_1 for $\boldsymbol{\rho}_1$ and consider the random draw of $\boldsymbol{\rho}_2$. The fact that the distribution of $\boldsymbol{\rho}_2$ is dependent on the actual values of ρ^{i-1} and ρ_1 is left implicit.

A set $\mathcal{F}$ of restrictions is said to be "downward closed" if changing the value of some coordinate of $\rho \in \mathcal{F}$ from the value $*$ to an element of $\{0,1\}$ cannot make it leave the set. Let us write this formally.

Definition 9.1. A set $\mathcal{F}$ of restrictions is *downward closed* if when $\rho \in \mathcal{F}$ and $\rho'(x_w) = \rho(x_w)$ for $w \neq w_0$ and $\rho(x_{w_0}) = *$ then $\rho' \in \mathcal{F}$.

We can now formulate the main lemma.

LEMMA 9.2. *Fix $2 \leq i < d$. Let $\boldsymbol{\rho}^i \leftarrow \mathcal{R}^i$ be a random restriction with components ρ^{i-1}, ρ_1, $\boldsymbol{\rho}_2$ and π (note that as discussed earlier we condition on arbitrary fixed values for ρ^{i-1} and ρ_1 so the only randomness is over $\boldsymbol{\rho}_2$). Let f be an arbitrary function. Suppose $g = f \restriction \rho^{i-1}$ is computed by a depth-2 circuit of bottom fan-in $t \leq 2^m/8$. Let $\mathcal{F}$ be a downward closed set of restrictions. Then, for sufficiently large m,*

$$\mathbf{Pr}_{\boldsymbol{\rho}^i \leftarrow \mathcal{R}^i}[\mathsf{depth}(g \restriction \boldsymbol{\rho}^i) \geq s \mid \boldsymbol{\rho}_2 \in \mathcal{F}] \leq D^s,$$

where $D = t2^{3-m/2}$.

PROOF. By symmetry we may assume that i is odd; recall that this means that each gate at level i of Sipser_d is an AND gate. By possibly looking at the negation of g (which has the same depth decision tree as g) we can assume that g is a CNF, so after ρ_1 has been applied it can be written as

$$g \restriction \rho_1 = \bigwedge_{i=1}^{\ell} C_i,$$

where each C_i is a disjunction of at most t literals.

The proof proceeds by induction over ℓ and the base case is when $\ell = 0$ in which case $g \restriction \boldsymbol{\rho}^i$ is always computable by a decision tree of depth 0.

For the inductive step, we divide the analysis into two cases depending on whether C_1 is forced to one or not. We can bound the probability of the lemma as the maximum of

$$\mathbf{Pr}_{\boldsymbol{\rho}^i}[\mathsf{depth}(g \restriction \boldsymbol{\rho}^i) \geq s \mid \boldsymbol{\rho}_2 \in \mathcal{F} \text{ and } C_1 \restriction \boldsymbol{\rho}_2 \equiv 1],$$

and

$$\mathbf{Pr}_{\boldsymbol{\rho}^i}[\mathsf{depth}(g \restriction \boldsymbol{\rho}^i) \geq s \mid \boldsymbol{\rho}_2 \in \mathcal{F} \text{ and } C_1 \restriction \boldsymbol{\rho}_2 \not\equiv 1]. \tag{10}$$

The first term is at most D^s as desired by induction applied to g without its first conjunction (note that this is a CNF with at most $\ell - 1$ clauses) and using that the conditioning in this case is a new downward closed set. In the following we analyze the second term (10).

In order for $g \restriction \boldsymbol{\rho}^i$ not to be identically 0 there must be some non-empty set Y of variables appearing in C_1 that are given the value $*$ by $\boldsymbol{\rho}_2$. For v a gate at level i of Sipser_d, let the set $\mathsf{Inputs}(\mathsf{Sipser}_v)$ of variables be called a "block," and let us suppose that the variables in C_1 come from t_1 different blocks. We say that a block is "undetermined by $\boldsymbol{\rho}_2$" if it contains a variable that is given the value $*$ by $\boldsymbol{\rho}_2$. For a set Z of blocks let us write $\mathrm{size}(Z)$ to denote the number of blocks it contains. We introduce the notation $\mathsf{undet}(Z, \boldsymbol{\rho}_2)$ to denote the event that Z is precisely the set of blocks that are undetermined by $\boldsymbol{\rho}_2$, and the notation $\det(\boldsymbol{\rho}_2, C_1 \setminus Z)$ to denote the event that all variables in C_1 outside of Z are fixed to non-$*$ values by $\boldsymbol{\rho}_2$.

We start constructing a decision tree for $g \restriction \boldsymbol{\rho}^i$ by querying the new variables in the blocks that are undetermined by $\boldsymbol{\rho}_2$. Let τ be an assignment to these variables. We can now bound (10) as

$$\sum_{\tau,Z} \mathbf{Pr}_{\boldsymbol{\rho}^i}[\mathrm{depth}(g \restriction \tau\boldsymbol{\rho}^i) \geq s - \mathrm{size}(Z) \text{ and } \mathsf{undet}(Z, \boldsymbol{\rho}_2) \text{ and } \det(\boldsymbol{\rho}_2, C_1 \setminus Z) \mid \boldsymbol{\rho}_2 \in \mathcal{F} \text{ and } C_1 \restriction \boldsymbol{\rho}_2 \not\equiv 1], \tag{11}$$

where in the sum Z ranges over all nonempty sets of blocks and τ ranges over all possible assignments to the variables in Z. Observe that because of the final projection step π in $\boldsymbol{\rho}^i$, each block in Z contributes only one to the decision tree depth.

We will use the upper bound

$$\mathbf{Pr}_{\boldsymbol{\rho}^i}[\mathrm{depth}(g \restriction \tau\boldsymbol{\rho}^i) \geq s - \mathrm{size}(Z) \mid \mathsf{undet}(Z, \boldsymbol{\rho}_2) \text{ and } \det(\boldsymbol{\rho}_2, C_1 \setminus Z) \text{ and } \boldsymbol{\rho}_2 \in \mathcal{F} \text{ and } C_1 \restriction \boldsymbol{\rho}_2 \not\equiv 1] \times \mathbf{Pr}_{\boldsymbol{\rho}^i}[\mathsf{undet}(Z, \boldsymbol{\rho}_2) \mid \boldsymbol{\rho}_2 \in \mathcal{F} \text{ and } C_1 \restriction \boldsymbol{\rho}_2 \not\equiv 1] \tag{12}$$

for each term in (11), and hence a key lemma is the following.

LEMMA 9.3. *If Z is a set of blocks appearing in C_1, then for $i \geq 2$ and sufficiently large m, we have*

$$\mathbf{Pr}_{\boldsymbol{\rho}^i}[\mathsf{undet}(Z, \boldsymbol{\rho}_2) \mid \boldsymbol{\rho}_2 \in \mathcal{F} \textit{ and } C_1 \restriction \boldsymbol{\rho}_2 \not\equiv 1] \leq 2^{\mathrm{size}(Z)(1-m/2)}. \tag{13}$$

PROOF. The crux of the proof is to, given an outcome ρ_2 of $\boldsymbol{\rho}_2$ that contributes to the probability in question, create a restriction ρ_2' that also satisfies the conditioning but fixes all variables in the blocks of Z. This easily gives (13) as argued at the end of this proof. We describe how to do this for the case when $\mathrm{size}(Z) = 1$ but the general case follows immediately as we can do the changes independently on each block. Thus let us assume that Z contains the single block $\mathsf{Inputs}(\mathsf{Sipser}_v)$ and fix an outcome ρ_2 of $\boldsymbol{\rho}_2$ that contributes to the event of the lemma, so $\mathsf{Inputs}(\mathsf{Sipser}_v)$ contains a variable that is given the value $*$ by ρ_2. Let P be the set of variables of $\mathsf{Inputs}(\mathsf{Sipser}_v)$ that appears positively in C_1 and N be the set of variables that appear negatively.

We can assume that we have no hard zero in $\mathsf{Inputs}(\mathsf{Sipser}_v)$ and that the number of non-hard ones in $\mathsf{Inputs}(\mathsf{Sipser}_v)$ is close to $f_i 2^{-m}$ (recall (2)^i) as otherwise already ρ_1 would have fixed all variables in $\mathsf{Inputs}(\mathsf{Sipser}_v)$ to constants.

Clearly we must have $\rho_2(x_v) = *$.[9] For variables $x_w \in P$ we must have $\rho_2(x_w) = *$ (recalling that there is no hard zero in $\mathsf{Inputs}(\mathsf{Sipser}_v)$) while for variables in N we have either $\rho_2(x_w) = *$ or $\rho_2(x_w) = 1$.

We now define a companion outcome $\rho_2' = H(\rho_2)$ for ρ_2. If ρ_2 maps some variable outside N to $*$ (and in particular if P is non-empty) we set $\rho_2'(x_v) = 0$ and otherwise $\rho_2'(x_v) = 1$. For each $x_w \in P$ we set $\rho_2'(x_w) = 0$ while for each $x_w \in N$ we set $\rho_2'(x_w) = 1$, independently of the value of $\rho_2(x_w)$. Outside C_1 but in $\mathsf{Inputs}(\mathsf{Sipser}_v)$ we set $\rho_2'(x_w) = 1$

[9] **Rocco:** Doesn't $\boldsymbol{\rho}_2$ act on variables at level $i-1$ like x_w, not x_v?

if $\rho_2(x_w) = 1$ and $\rho_2'(x_w) = \rho_2'(x_v)$ otherwise. Outside $\mathsf{Inputs}(\mathsf{Sipser}_v)$, ρ_2 and ρ_2' agree. First observe that ρ_2' satisfies the conditioning, since we only have $\rho_2(x_w) \neq \rho_2'(x_w)$ when $\rho_2(x_w) = *$ and by the definition of P and N we are careful not to satisfy C_1.

The mapping H is many-to-one as given ρ_2' we do not know the values of $\rho_2(x_w)$ when $x_w \in N$ (but we do for all other variables in $\mathsf{Inputs}(\mathsf{Sipser}_v)$).

We note that we have

$$\frac{\mathbf{Pr}_{\boldsymbol{\rho}^i}[\boldsymbol{\rho}_2 = \rho_2]}{\mathbf{Pr}_{\boldsymbol{\rho}^i}[\boldsymbol{\rho}_2 = \rho_2']} = \frac{\mathbf{Pr}_{\boldsymbol{\rho}^i}[\boldsymbol{\rho}_{2,v} = \rho_{2,v}]}{\mathbf{Pr}_{\boldsymbol{\rho}^i}[\boldsymbol{\rho}_{2,v} = \rho_{2,v}']}$$

where $\boldsymbol{\rho}_{2,v}$ is only the behavior of $\boldsymbol{\rho}_2$ on $\mathsf{Inputs}(\mathsf{Sipser}_v)$ and similarly for $\rho_{2,v}'$. This is true as ρ_2 and ρ_2' take the same values outside $\mathsf{Inputs}(\mathsf{Sipser}_v)$ and the restrictions are picked independently on each block.

Assume first that $\rho_2'(x_v) = 0$. In this situation ρ_2 could have been picked under case $3^{(i)}$ or case $4^{(i)}$ while ρ_2' can only have been produced under case $4^{(i)}$. We know, by (9), that each ρ_2 is about a factor 2^m more likely to have been produced under case $4^{(i)}$ than under case $3^{(i)}$ so let us ignore case $3^{(i)}$, introducing a small error factor $(1 + O(2^{-m}))$ that we temporarily suppress.

Let N_1 denote the subset of N that was actually given the value $*$ by $\boldsymbol{\rho}_2$. If N_1 is empty then $\mathbf{Pr}_{\boldsymbol{\rho}^i}[\boldsymbol{\rho}_{2,v} = \rho_{2,v}] = \frac{q_4(\boldsymbol{f}_v(\boldsymbol{\alpha}))}{1-q_4(\boldsymbol{f}_v(\boldsymbol{\alpha}))} \mathbf{Pr}_{\boldsymbol{\rho}^i}[\boldsymbol{\rho}_{2,v} = \rho_{2,v}']$ and in general we pick up an additional factor of $b_{i-1}^{|N_1|}(1-b_{i-1})^{-|N_1|}$. As

$$\sum_{N_1 \subseteq N} b_{i-1}^{|N_1|}(1-b_{i-1})^{-|N_1|} = (1 + \frac{b_{i-1}}{1-b_{i-1}})^{|N|}$$

we get

$$\begin{aligned}\sum_{H(\rho_2)=\rho_2'} \mathbf{Pr}_{\boldsymbol{\rho}^i}[\boldsymbol{\rho}_2 = \rho_2] &\leq \left(1 + \frac{b_{i-1}}{1-b_{i-1}}\right)^{|N|} \frac{q_4(\boldsymbol{f}_v(\boldsymbol{\alpha}))}{1-q_4(\boldsymbol{f}_v(\boldsymbol{\alpha}))} \mathbf{Pr}_{\boldsymbol{\rho}^i}[\boldsymbol{\rho}_2 = \rho_2'] \\ &\leq 2^{1-m} \mathbf{Pr}_{\boldsymbol{\rho}^i}[\boldsymbol{\rho}_2 = \rho_2'] \end{aligned} \tag{14}$$

for sufficiently large m. This follows as $|N| \leq 2^m/8$, $b_i = (1+o(1))2^{-m}$ and $q_4(\boldsymbol{f}_v(\boldsymbol{\alpha})) = (1+o(1))2^{-m}$.

If, $\rho_2'(x_v) = 1$ the situation is similar except that ρ_2' is produced under case $3^{(i)}$ (while we still only consider the ρ_2 produced under case $4^{(i)}$) and thus we pick up a factor q_3 instead of $1 - q_4(\boldsymbol{f}_v(\boldsymbol{\alpha}))$. We get in this case

$$\begin{aligned}\sum_{H(\rho_2)=\rho_2'} \mathbf{Pr}_{\boldsymbol{\rho}^i}[\boldsymbol{\rho}_2 = \rho_2] &\leq \left(1 + \frac{b_{i-1}}{1-b_{i-1}}\right)^{|N|} \frac{q_4(\boldsymbol{f}_v(\boldsymbol{\alpha}))}{q_3} \mathbf{Pr}_{\boldsymbol{\rho}^i}[\boldsymbol{\rho}_2 = \rho_2'] \\ &\leq 2^{1-m/2} \mathbf{Pr}_{\boldsymbol{\rho}^i}[\boldsymbol{\rho}_2 = \rho_2'], \end{aligned} \tag{15}$$

again for sufficiently large m. The fact that we ignored restrictions ρ_2 produced under case $3^{(i)}$ gives an additional factor $(1 + O(2^{-m}))$ in the above estimates and thus the calculations remain valid, possibly by making m slightly larger to make sure that the "sufficiently large m" statements are true.

As mentioned earlier, the case of general $\mathrm{size}(Z) > 1$ follows from the fact that we do the modifications on all blocks of Z independently. □

Remark 9.4. The careful reader might have noticed that in the case when $\rho_2'(x_v) = 1$ we can infer that N_1 is non-empty, giving a slightly better estimate especially in the case when t is small. This observation can probably be used to get a slightly better

constant in the main theorem, but to keep the argument simple we ignore this point. We return to the main argument.

We now upper bound

$$\mathbf{Pr}_{\boldsymbol{\rho}^i}[\mathsf{depth}(g \restriction \tau\boldsymbol{\rho}^i) \geq s - \mathrm{size}(Z) \mid \mathsf{undet}(Z, \boldsymbol{\rho}_2) \text{ and } \det(\boldsymbol{\rho}_2, C_1 \setminus Z) \text{ and } \boldsymbol{\rho}_2 \in \mathcal{F} \text{ and } C_1 \restriction \boldsymbol{\rho}_2 \not\equiv 1] \tag{16}$$

by $D^{s-\mathrm{size}(Z)}$ using the inductive hypothesis. We need to check that the conditioning defines a downward closed set. This is not complicated but let us give some details. Fix any behavior of $\boldsymbol{\rho}_2$ inside the blocks of Z and satisfying the conditioning. As $g \restriction \tau\boldsymbol{\rho}^i$ does not depend on the variables corresponding to Z, the event in (16) depends only on the values of $\boldsymbol{\rho}_2$ outside Z. Changing $\boldsymbol{\rho}_2$ from $*$ to a constant value for any variable outside Z cannot violate any of the conditions in the conditioning and hence we have a downward closed set when considering $\boldsymbol{\rho}_2$ as a restriction outside Z. We conclude that the probability of the event in (16) is, by induction, bounded by $D^{s-\mathrm{size}(Z)}$.

Recall that our goal is to upper bound the sum (11), using the bound (12) for each term, Lemma 9.3, and the inductive hypothesis. If C_1 intersects t_1 different blocks (where of course $t_1 \leq t$) then, using the fact that we have at most $2^{\mathrm{size}(Z)}$ different τ and recalling the setting $D = t2^{3-m/2}$ we get the total estimate

$$\sum_{Z \neq \emptyset} 2^{\mathrm{size}(Z)} 2^{\mathrm{size}(Z)(1-m/2)} D^{s-\mathrm{size}(Z)} = D^s \left((1 + D^{-1}2^{2-m/2})^{t_1} - 1 \right) \leq D^s \left((1 + \frac{1}{2t})^{t_1} - 1 \right) \leq D^s$$

and we are done with the proof of Lemma 9.2.

Lemma 9.2 is sufficient to prove a fairly tight hierarchy theorem. To prove a tight variant we need also to analyze how $\mathcal{R}^1$ simplifies circuits.

LEMMA 9.5. *Let g be computed by a depth-2 circuit of bottom fan-in $t \leq m/4$. Let $\mathcal{F}$ be a downward closed set of restrictions and consider a random projection $\boldsymbol{\rho}^1 \leftarrow \mathcal{R}^1$. For sufficiently large m,*

$$\mathbf{Pr}[\mathsf{depth}(g \restriction \boldsymbol{\rho}^1) \geq s \mid \boldsymbol{\rho} \in \mathcal{F}] \leq D^s,$$

where $D = t2^{3+t-m/2}$.

PROOF. The proof of this lemma is almost identical to the proof of Lemma 9.2 so we only discuss the differences. Lemma 9.3 is replaced by the following.

LEMMA 9.6. *If Z is a set of blocks appearing in C_1 and $\boldsymbol{\rho}$ is a random restriction appearing in the construction of $\boldsymbol{\rho}^1 \leftarrow \mathcal{R}^1$ (see Definition 7.2), then, for sufficiently large m,*

$$\mathbf{Pr}_{\boldsymbol{\rho}}[\mathsf{undet}(Z, \boldsymbol{\rho}) \mid \boldsymbol{\rho} \in \mathcal{F} \text{ and } C_1 \restriction \boldsymbol{\rho} \not\equiv 1] \leq 2^{r(t+1-m/2)}.$$

PROOF. The proof is almost the same as the proof of Lemma 9.3. The reason for the loss in parameters is that the factor

$$\left(1 + \frac{b_{i-1}}{1 - b_{i-1}}\right)^{|N|}$$

that used to be bounded by a constant strictly less than two can now be as large as 2^t. □

The rest of the proof of how Lemma 9.5 follows from Lemma 9.6 is identical with how Lemma 9.2 followed from Lemma 9.3 with the obvious change in the final calculation.

10. THE PROOF OF THE MAIN THEOREMS

We now proceed to prove Theorem 1.6. In fact we are going to prove the following, slightly stronger, theorem.

THEOREM 10.1. *Let C be a circuit of depth d and size S with bottom fan-in at most $m/4$. Then for sufficiently large m, we have*

$$\mathbf{Pr}[\mathsf{Sipser}_d(x) = C(x)] \leq \frac{1}{2} + O(2^{-m/4}) + S2^{-2^{m/2-4}}.$$

Recalling that $m = \frac{\log n}{2d-2}(1 + o(1))$, it is not difficult to see that this theorem implies Theorem 1.6. We turn to proving Theorem 10.1.

PROOF. We analyze what happens when a random projection $\boldsymbol{\rho}^{d-1} \in \mathcal{R}^{d-1}$ is applied to both Sipser_d and C. Let us assume that d is odd and hence the output gate of Sipser_d is an AND gate. (The case of even d is completely analogous.) By Lemma 8.4 we have

$$\mathbf{Pr}_{\boldsymbol{x}\leftarrow\{0,1\}^n}[\mathsf{Sipser}_d(\boldsymbol{x}) = C(\boldsymbol{x})] = \mathbf{Pr}_{\boldsymbol{x}\leftarrow\{0,1\}^{V_{d-1}},\boldsymbol{\rho}^{d-1}\leftarrow\mathcal{R}^{d-1}}[\mathsf{Sipser}_d \restriction \rho^{d-1}(\boldsymbol{x}) = C \restriction \rho^{d-1}(\boldsymbol{x})]$$

where we stress that on the LHS the probability is over a uniform random n-bit string $\boldsymbol{x}$ while on the RHS the probability is over a random draw of $\boldsymbol{\rho}^{d-1}$ from $\mathcal{R}^{d-1}$ and a coordinate-wise independent assignment to the live variables in V_{d-1} where each variable is given the value 1 with probability $1 - b_{d-1}$. Let us first see how $\boldsymbol{\rho}^{d-1}$ affects Sipser_d.

Consider the output gate v of Sipser_d, which has fan-in f_d. Recalling key property (A) and the value of f_d from (2) (and observing that each input gate to v is at level $d-1$ which is even), we get that with probability $O(2^{-m/2})$ some gate that is an input to v is forced to 0 by $\boldsymbol{\rho}^{d-1}$. Suppose that this does not happen and let $\boldsymbol{h}_1$ denote the number of input gates to v that are not fixed to one. By key property (B) and a Chernoff bound applied to each input gate to v and a union bound over those input gates, with probability $1 - \exp(-\Omega(2^{m/2}))$ we have $|\boldsymbol{h}_1 - f_d 2^{-m}| \leq 2^{3m/4}$. Thus we conclude that, with probability $1 - O(2^{-m/2})$, $\mathsf{Sipser}_d \restriction \boldsymbol{\rho}^{d-1}$ has been reduced to an AND-gate of fan-in $(\ln 2) \cdot 2^m \cdot (1 \pm O(2^{-m/4}))$.

Now let us see how $\boldsymbol{\rho}^{d-1}$ affects C. We aim to prove by induction that $\boldsymbol{\rho}^i \leftarrow \mathcal{R}^i$, with high probability, reduces the depth of C by i. Let us assume that C has S_i gates at distance i from the inputs.

Consider any gate in C at distance two from the inputs and suppose it is an OR gate whose inputs are AND gates, the other case being similar. By Lemma 9.5, for sufficiently large m, after $\boldsymbol{\rho}^1 \leftarrow \mathcal{R}^1$ has been applied, except with probability $2^{-2^{m/2-4}}$ this sub-circuit can be computed by a decision tree of depth at most $2^{m/2-4}$. This implies that it can be written as an AND of OR gates of fan-in at most $2^{m/2-4}$. We conclude that except with probability at most $S_2 \cdot 2^{-2^{m/2-4}}$, by collapsing two adjacent levels of AND-gates, $C \restriction \boldsymbol{\rho}^1$ can be computed by a depth $d-1$ circuit with bottom fan-in at most $2^{m/2-4}$ where each gate at distance at least two from the inputs corresponds to a gate at distance at least three in the original circuit.

Applying Lemma 9.2 for $i = 2, 3 \ldots d-2$ in a similar way we conclude that for a random draw of $\boldsymbol{\rho}^2 \leftarrow \mathcal{R}^{d-2}$, except with probability at most $\sum_{i=2}^{d-2} S_i 2^{-2^{m/2-4}}$, $C \restriction \boldsymbol{\rho}^{d-2}$ can be computed by a depth 2 circuit of bottom fan-in $2^{m/2-4}$. A final application of Lemma 9.2 says that except with an additional failure probability $2^{-2^{m/2-4}}$, for a random draw of $\boldsymbol{\rho}^{d-1} \leftarrow \mathcal{R}^{d-1}$, $C \restriction \boldsymbol{\rho}^{d-1}$ can be computed by a decision tree of depth $2^{m/2-4}$.

Summarizing the above analysis, we have that except with probability $O(2^{-m/2}) + S2^{-2^{m/2-4}}$ over a random draw of $\boldsymbol{\rho}^{d-1} \leftarrow \mathcal{R}^{d-1}$, it is true both that $\mathsf{Sipser}_d \restriction \boldsymbol{\rho}^{d-1}$ is an

AND of fan-in $\ln 2 \cdot 2^m(1 \pm O(2^{-m/4}))$ and that $C \restriction \boldsymbol{\rho}^{d-1}$ is computed by a decision tree of depth $2^{m/2-4}$. For coordinate-wise independent input strings where each variable is given the value 1 with probability $1 - b_{d-1}$, an AND of fan-in $\ln 2 \cdot 2^m(1 \pm O(2^{-m/4}))$ evaluates to 1 with probability $\frac{1}{2}(1 \pm O(2^{-m/4}))$ (recalling (4)), while the probability that any decision tree of depth s outputs 1 must be within sb_{d-1} of either 0 or 1. We conclude that

$$\mathbf{Pr}_{\boldsymbol{x} \leftarrow \{0,1\}^{V_{d-1}}, \boldsymbol{\rho}^{d-1} \leftarrow \mathcal{R}^{d-1}}[\mathsf{Sipser}_d \restriction \rho^{d-1}(\boldsymbol{x}) = C \restriction \rho^{d-1}(\boldsymbol{x})] = \frac{1}{2} \pm \left(O(2^{-m/4}) + S2^{-2^{m/2-4}} \right),$$

and the proof is complete. □

Looking more closely at the proof, we can derive an even stronger theorem that implies Theorem 1.7:

THEOREM 10.2. *For odd d, let C be a circuit of size at most S and depth $d+1$ with an* OR*-gate as its output gate, and bottom fan-in at most $m/4$. Then for sufficiently large m we have*

$$\mathbf{Pr}_{\boldsymbol{x} \leftarrow \{0,1\}^n}[\mathsf{Sipser}_d(\boldsymbol{x}) = C(\boldsymbol{x})] \leq \frac{1}{2} + O(2^{-m/4}) + S2^{-2^{m/2-4}}.$$

The same is true for even d if the output gate of C is an AND*-gate.*

PROOF. Let us assume that d is odd, the even case being completely analogous. We follow exactly the proof of Theorem 10.1 until the very last step. We can conclude that $C \restriction \boldsymbol{\rho}^{d-1}$, with high probability, is reduced to the disjunction of a set of functions each computable by a decision tree of depth $2^{m/2-4}$. We can convert this to a DNF formula of bottom fan-in $2^{m/2-4}$, and we must analyze the probability that such a DNF equals $\mathsf{Sipser}_d \restriction \boldsymbol{\rho}^{d-1}$ under the coordinate-wise independent distribution on input strings where each variable is given the value 1 with probability $1 - b_{d-1}$. There are two cases to consider.

Suppose first that each term in the DNF formula contains a negated variable. Then $C \restriction \boldsymbol{\rho}^{d-1}$ rejects the all-ones input, which is chosen with probability $\frac{1}{2} + O(2^{-m/4})$. On the other hand, if $\mathsf{Sipser}_d \restriction \boldsymbol{\rho}^{d-1}$ is an AND-gate of fan-in $\ln 2 \cdot 2^m(1 \pm O(2^{-m/4}))$ (which happens with probability $1 - O(2^{-m/2})$), then the all-ones input is accepted by $\mathsf{Sipser}_d \restriction \boldsymbol{\rho}^{d-1}$ with probability 1. We thus have that

$$\mathbf{Pr}_{\boldsymbol{\rho}^{d-1}}[\mathsf{Sipser}_d \restriction \rho^{d-1}(\boldsymbol{x}) = C \restriction \rho^{d-1}(\boldsymbol{x})] \leq \frac{1}{2} + O(2^{-m/2}) + O(2^{-m/4}) \tag{17}$$

in this case, giving the desired bound.

On the other hand if there is a term in $C \restriction \boldsymbol{\rho}^{d-1}$ that only contains positive variables then it (and hence $C \restriction \boldsymbol{\rho}^{d-1}$) is true with probability $1 - O(2^{-m/2})$. As $\mathsf{Sipser}_d \restriction \boldsymbol{\rho}^{d-1}$ is close to unbiased, (17) is true also in this case and the theorem follows. □

As stated previously we have not made any serious effort to get the best constants in our main theorems. Our constants are, however, not too far from the truth as we may take C to be one input sub-circuit to the output gate of Sipser_d. This is a depth $d-1$ circuit of sub-linear size that agrees with Sipser_d for a fraction $\frac{1}{2} + \Omega(2^{-2m})$ of all inputs.

11. SOME FINAL WORDS

The main difference between the current paper and the early proof of the hierarchy theorem in [Håstad 1986a] is the use of projections. The projections serve two purposes. The first is to make sure that once a single $*$ is found in ρ we do not bias any other value of ρ^i to be $*$. This property was achieved in [Håstad 1986a] by fixing the

values of neighboring variables to constants while here we identify all the neighboring variables with the same new variable and hence we only query one variable in the decision tree. We feel that this difference is minor.

The more important difference is that projections enables us to choose a uniformly random input where this seemed difficult to achieve. It is remarkable how seemingly simple ideas can take care of problems that, at least initially, looks like fundamental obstacles.

REFERENCES

Scott Aaronson. The Complexity Zoo. (????). Available at http://cse.unl.edu/~cbourke/latex/ComplexityZoo.pdf.

Scott Aaronson. 2010a. A Counterexample to the Generalized Linial-Nisan Conjecture. *Electronic Colloquium on Computational Complexity* 17 (2010), 109.

Scott Aaronson. 2010b. BQP and the polynomial hierarchy. In *Proceedings of the 42nd ACM Symposium on Theory of Computing*. 141–150.

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

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

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

László Babai. 1987. Random oracles separate PSPACE from the polynomial-time hierarchy. *Inform. Process. Lett.* 26, 1 (1987), 51–53.

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

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

Louay Bazzi. 2009. Polylogarithmic independence can fool DNF formulas. *SIAM J. Comput.* 38, 6 (2009), 2220–2272.

Paul Beame, Russell Impagliazzo, and Srikanth Srinivasan. 2012. 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*. 117–125.

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

Charles Bennett and John Gill. 1981. Relative to a Random Oracle A, $\mathsf{P}^A \neq \mathsf{NP}^A \neq \mathsf{coNP}^A$ with Probability 1. *SIAM J. on Comput.* 10, 1 (1981), 96–113.

Ronald Book. 1994. On collapsing the polynomial-time hierarchy. *Inform. Process. Lett.* 52, 5 (1994), 235–237.

Ravi Boppana. 1997. The Average Sensitivity of Bounded-Depth Circuits. *Inform. Process. Lett.* 63, 5 (1997), 257–261.

Mark Braverman. 2010. Polylogarithmic independence fools AC^0 circuits. *J. ACM* 57, 5 (2010), 28.

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

Jin-Yi Cai. 1986. 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*. 21–29.

Liming Cai, Jianer Chen, and Johan Håstad. 1998. Circuit bottom fan-in and computational power. *SIAM J. Comput.* 27, 2 (1998), 341–355.

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

Lance Fortnow. 1999. Relativized Worlds with an Infinite Hierarchy. *Inform. Process. Lett.* 69, 6 (1999), 309–313.

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

Oded Goldreich and Avi Wigderson. 2013. On the Size of Depth-Three Boolean Circuits for Computing Multilinear Functions. *Electronic Colloquium on Computational Complexity* (2013).

András Hajnal, Wolfgang Maass, Pavel Pudlák, Mário Szegedy, and György Turán. 1993. Threshold circuits of bounded depth. *J. Comput. System Sci.* 46 (1993), 129–154.

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

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

Johan Håstad. 1989. *Almost optimal lower bounds for small depth circuits*. JAI Press, 143–170.

Johan Håstad. 2014. On the Correlation of Parity and Small-Depth Circuits. *SIAM J. Comput.* 43, 5 (2014), 1699–1708.

J. Håstad. 2016. An Average-case Depth Hierarchy Theorem for Higher Depths. In *Proc. 57th IEEE FOCS*.

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

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

Lane Hemaspaandra and Mitsunori Ogihara. 2002. *The Complexity Theory Companion*. Springer.

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

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

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

David Johnson. 1986. The NP-Completeness Column: An Ongoing Guide. *Journal of Algorithms* 7, 2 (1986), 289–305.

Stasys Jukna. 2012. *Boolean Function Complexity*. Springer.

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

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

Nathan Linial, Yishay Mansour, and Noam Nisan. 1993. Constant Depth Circuits, Fourier Transform, and Learnability. *J. ACM* 40, 3 (1993), 607–620.

Yishay Mansour. 1995. An $O(n^{\log\log n})$ learning algorithm for DNF under the uniform distribution. *J. Comput. System Sci.* 50 (1995), 543–550.

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

Ryan O'Donnell. 2007. 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.

Ryan O'Donnell and Karl Wimmer. 2007. Approximation by DNF: Examples and Counterexamples. In *34th International Colloquium on Automata, Languages and Programming*. 195–206.

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

Alexander Razborov. 1987. 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 (1987), 333–338.

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

Benjamin Rossman. 2015. The Average Sensitivity of Bounded-Depth Formulas. In *Proceedings of the 56th Annual Symposium on Foundations of Computer Science*. 424–430.

Benjamin Rossman, Rocco A. Servedio, and Li-Yang Tan. 2015a. Complexity Theory Column 89: The Polynomial Hierarchy, Random Oracles, and Boolean Circuits. *SIGACT News* 46, 4 (2015), 50–68.

Benjamin Rossman, Rocco A Servedio, and Li-Yang Tan. 2015b. An average-case depth hierarchy theorem for Boolean circuits. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*. IEEE, 1030–1048.

Nathan Segerlind, Sam Buss, and Russell Impagliazzo. 2004. A switching lemma for small restrictions and lower bounds for k-DNF resolution. *SIAM J. Comput.* 33, 5 (2004), 1171–1200.

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

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

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

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

Gábor Tardos. 1989. 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 (1989), 385–392.

Leslie Valiant. 1983. Exponential Lower Bounds for Restricted Monotone Circuits. In *Proceedings of the 15th Annual ACM Symposium on Theory of Computing*. 110–117.

Emanuele Viola. 2013. Challenges in computational lower bounds. *Electronic Colloquium on Computational Complexity* (2013).

Heribert Vollmer and Klaus Wagner. 1997. *Measure One Results in Computational Complexity Theory*. Springer, 285–312.

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

Ryan Williams. 2014b. 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*.

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

Received (month) (year); revised (month) (year); accepted (month) (year)