A Technique for Hardness Amplification Against AC^0

William M. Hoza

Department of Computer Science, The University of Chicago, IL, USA

Abstract

We study hardness amplification in the context of two well-known "moderate" average-case hardness results for AC^0 circuits. First, we investigate the extent to which AC^0 circuits of depth d can approximate AC^0 circuits of some larger depth $d+k$. The case $k=1$ is resolved by Håstad, Rossman, Servedio, and Tan's celebrated average-case depth hierarchy theorem (JACM 2017). Our contribution is a significantly stronger correlation bound when $k \geq 3$. Specifically, we show that there exists a linear-size AC^0_{d+k} circuit $h\colon \{0,1\}^n \to \{0,1\}$ such that for every AC^0_d circuit g, either g has size $\exp(n^{\Omega(1/d)})$, or else g agrees with h on at most a $(1/2+\varepsilon)$-fraction of inputs where $\varepsilon = \exp(-(1/d)\cdot\Omega(\log n)^{k-1})$. For comparison, Håstad, Rossman, Servedio, and Tan's result has $\varepsilon = n^{-\Theta(1/d)}$. Second, we consider the majority function. It is well known that the majority function is moderately hard for AC^0 circuits (and stronger classes). Our contribution is a stronger correlation bound for the XOR of t copies of the n-bit majority function, denoted $\mathsf{MAJ}_n^{\oplus t}$. We show that if g is an AC^0_d circuit of size S, then g agrees with $\mathsf{MAJ}_n^{\oplus t}$ on at most a $(1/2+\varepsilon)$-fraction of inputs, where $\varepsilon = \left(O(\log S)^{d-1}/\sqrt{n}\right)^t$.

To prove these results, we develop a hardness amplification technique that is tailored to a specific type of circuit lower bound proof. In particular, one way to show that a function h is moderately hard for AC^0 circuits is to (a) design some distribution over random restrictions or random projections, (b) show that AC^0 circuits simplify to shallow decision trees under these restrictions/projections, and finally (c) show that after applying the restriction/projection, h is moderately hard for shallow decision trees with respect to an appropriate distribution. We show that (roughly speaking) if h can be proven to be moderately hard by a proof with that structure, then XORing multiple copies of h amplifies its hardness. Our analysis involves a new kind of XOR lemma for decision trees, which might be of independent interest.

2012 ACM Subject Classification Theory of computation → Circuit complexity

Keywords and phrases Bounded-depth circuits, average-case lower bounds, hardness amplification, XOR lemmas

Digital Object Identifier 10.4230/LIPIcs.CCC.2024.1

Related Version *Full Version*: https://eccc.weizmann.ac.il/report/2023/176/ [41]

Funding Part of this work was done while the author was visiting the Simons Institute for the Theory of Computing.

Acknowledgements I thank Avishay Tal for collaboration at an early stage of this project. I thank Li-Yang Tan for encouragement. I thank Pooya Hatami for a helpful conversation.

1 Introduction

1.1 Average-Case Circuit Lower Bounds

Circuit lower bounds are at the heart of computational complexity theory. To understand the limitations of (extremely) efficient computation, we seek to prove that certain explicit functions cannot be computed by certain interesting classes of Boolean circuits. In fact, ideally, we want to prove *average-case* circuit lower bounds, also known as *correlation bounds*. That is, we would like to prove that circuits in some class $\mathcal{C}$ cannot compute some function $h\colon \{0,1\}^n \to \{0,1\}$ on more than a $(1/2+\varepsilon)$-fraction of inputs for some small value $\varepsilon > 0$:

39th Computational Complexity Conference (CCC 2024).
Editor: Rahul Santhanam; Article No. 1; pp. 1:1–1:20
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

$$\text{For every } g \in \mathcal{C}, \quad \Pr_{\mathbf{x}\in\{0,1\}^n}[g(\mathbf{x}) = h(\mathbf{x})] \le \frac{1}{2} + \varepsilon. \tag{1}$$

We would like ε to be as small as possible. For example, one motivation for trying to minimize ε comes from the Nisan-Wigderson framework for converting correlation bounds into pseudorandom generators (PRGs) [57]. In this framework, a bound of the form (1) implies a PRG with error εn, and in particular, the framework requires $\varepsilon < 1/n$.

In this work, we focus on the case that $\mathcal{C}$ consists of AC^0 circuits, i.e., circuits made up of AND and OR gates of unbounded fan-in, with literals and constants at the bottom. The *size* of the circuit is the number of AND and OR gates, and the *depth* of the circuit is the length of the longest path from an input gate to the output gate. We refer to an AC^0 circuit of depth d as an "AC^0_d circuit." We are especially interested in the constant-depth regime; this class of circuits can be viewed as a model of constant-time parallel computation. Some of the most celebrated theorems in circuit complexity are lower bounds on the size of AC^0 circuits computing various explicit functions. For example, if g is an AC^0_d circuit, then g famously cannot compute the parity function on n bits or the majority function on n bits, unless g has size at least $\exp(c_d \cdot n^{1/(d-1)})$ [28, 1, 80, 35, 36].

1.2 Hardness Amplification and Yao's XOR Lemma

One appealing approach for proving strong correlation bounds is to first construct a function h that is "moderately hard" (e.g., maybe we have $\varepsilon = 1/\sqrt{n}$), and then apply some kind of *hardness amplification* scheme that converts h into a "very hard" function (e.g., maybe now we can take $\varepsilon = n^{-\omega(1)}$). The most famous method for hardness amplification is Yao's XOR Lemma [79, 53, 43, 29]. Starting from a hard function $h\colon \{0,1\}^n \to \{0,1\}$, this lemma considers the new hard function $h^{\oplus t}\colon \{0,1\}^{nt} \to \{0,1\}$ defined by $h^{\oplus t}(x^{(1)}, \dots, x^{(t)}) = \bigoplus_{i=1}^t h(x^{(i)})$. One well-known version[1] of Yao's XOR Lemma says that if h is moderately hard for $\mathsf{MAJ} \circ \mathcal{C}$ circuits, where MAJ denotes the majority function, then $h^{\oplus t}$ is very hard for $\mathcal{C}$ circuits.

In the context of relatively weak classes such as AC^0, the distinction between $\mathcal{C}$ and $\mathsf{MAJ} \circ \mathcal{C}$ is extremely important. Proving lower bounds on the size of $\mathsf{MAJ} \circ \mathcal{C}$ circuits is generally much more difficult than proving lower bounds on the size of $\mathcal{C}$ circuits. For this reason, there is a great deal of interest in "removing the majority gate" from Yao's XOR Lemma. For example, we can ask the following.

▶ **Question 1.1** (Does XORing amplify hardness for AC^0?). *Let $h\colon \{0,1\}^n \to \{0,1\}$ and let $t = \log n$. Assume that every constant-depth subexponential-size AC^0 circuit g satisfies*

$$\Pr_{\mathbf{x}\in\{0,1\}^n}[g(\mathbf{x}) = h(\mathbf{x})] \le \frac{1}{2} + n^{-\Omega(1)}.$$

Does it follow that every constant-depth polynomial-size AC^0 circuit g satisfies

$$\Pr_{\mathbf{x}\in\{0,1\}^{nt}}[g(\mathbf{x}) = h^{\oplus t}(\mathbf{x})] \le \frac{1}{2} + n^{-\omega(1)}\,?$$

Several recent papers have developed and applied a refined version of Yao's XOR Lemma featuring an "approximate linear sum" gate instead of the traditional majority gate [22, 21, 20, 42, 19, 25]. This clever approach has been fruitful, but it is still not applicable if we

[1] See, for example, Viola's work [77].

start with a function that is hard merely for AC^0 circuits. Unfortunately, there are strong *barrier results* saying that every "black-box" hardness amplification scheme must involve *some* nontrivial computational overhead [74, 32, 63, 31, 62]. As a special case, this line of work implies that Theorem 1.1 cannot be resolved affirmatively via a "black-box" hardness amplification scheme. Thus, we have an ironic state of affairs: we have a rich toolkit for proving lower bounds on the size of AC^0 circuits, because we are able to exploit these circuits' weaknesses, but at the same time, *specifically because these circuits are too weak*, we cannot use Yao's XOR Lemma to amplify our lower bounds.[2]

1.3 Our Contributions

In this work, we develop a non-black-box method for hardness amplification, applicable to some (but not all) moderate hardness results for AC^0 circuits. We use our method to amplify two well-known average-case hardness results, discussed next.

1.3.1 Correlation Bounds for Depth Reduction Within AC^0

Our first application of our hardness amplification technique concerns the role of depth in circuit complexity. To what extent are deeper circuits more powerful than shallower circuits? In other words, what is the *marginal utility of time* for parallel computation?

Surprisingly, it turns out that in many contexts, circuits can be generically and nontrivially simulated by shallower circuits. For example:

- NC^1 circuits (i.e., circuits of depth $O(\log n)$ with bounded fan-in) can be simulated by AC^0_d circuits of size $\exp(n^{O(1/d)})$ [73, 75, 76, 71].
- ACC^0_d circuits (i.e., AC^0_d circuits augmented with MOD_m gates) of size S can be simulated by $\mathsf{SYM} \circ \mathsf{AND}$ circuits of size $\exp((\log S)^{O(d)})$ [72, 2, 4, 81, 3, 11, 78, 24].
- AC^0 circuits can be approximated in various ways by low-degree polynomials [60, 66, 67, 10, 70, 55, 15, 37, 8, 59, 16, 69, 51, 34], which can be viewed as a "depth-two" model of computation.

In light of these remarkable "depth reduction" results and their numerous applications, we would like to know precisely when, and to what extent, depth reduction is possible. Indeed, there is a longstanding interest in thoroughly understanding the *hardness of circuit depth reduction* within AC^0. Early work shows that there exists a linear-size AC^0_{d+1} circuit $h \colon \{0,1\}^n \to \{0,1\}$ such that every AC^0_d circuit computing h must have size $\exp(n^{\Omega(1/d)})$ [65, 80, 35]. For several decades, it was a stubborn open problem to prove a similar hierarchy theorem in the average-case setting. O'Donnell and Wimmer essentially resolved the depth-2 vs. depth-3 case [58], and then finally Håstad, Rossman, Servedio, and Tan resolved the general depth-d vs. depth-$(d+1)$ case in a breakthrough last decade [39]:

▶ **Theorem 1.2** (The average-case depth hierarchy theorem [39]). *Let $n, d \in \mathbb{N}$ with $d \leq \frac{\alpha \log n}{\log \log n}$, where $\alpha > 0$ is a suitable constant. There is an explicit*[3] *AC^0_{d+1} circuit $h \colon \{0,1\}^n \to \{0,1\}$ of size $O(n)$ such that for every AC^0_d circuit $g \colon \{0,1\}^n \to \{0,1\}$, either g has size $\exp(n^{\Omega(1/d)})$, or else the following correlation bound holds:*

$$\Pr_{\mathbf{x} \in \{0,1\}^n}[g(\mathbf{x}) = h(\mathbf{x})] \leq \frac{1}{2} + n^{-\Omega(1/d)}. \tag{2}$$

[2] The exception, of course, is if we start from a lower bound against a stronger class such as $\mathsf{MAJ} \circ \mathsf{AC}^0$. See Klivans' work [49].

[3] I.e., the circuit h can be constructed in $\mathrm{poly}(n)$ time, given the parameters n and d.

Theorem 1.2 asserts that h is *moderately* hard for AC^0_d circuits. Håstad, Rossman, Servedio, and Tan identified two obstacles preventing significant improvement of the $n^{-\Omega(1/d)}$ correlation bound in (2):

- The "hard function" h in Theorem 1.2 is monotone. By the Kahn-Kalai-Linial theorem [47], every monotone Boolean function can be approximated by a constant or a variable with success probability $1/2 + \omega(1/n)$.
- By the discriminator lemma [33], every linear-size AC^0_{d+1} circuit h, whether monotone or not, can be approximated by a linear-size AC^0_d circuit with success probability $1/2+\Omega(1/n)$.

(See Hatami, Hoza, Tal, and Tell's work for further details of these two arguments [40, Appendix A].)

In this work, we overcome both obstacles by using a different, non-monotone hard function h with depth slightly greater than $d + 1$. We prove an average-case lower bound for the task of simulating AC^0_{d+k} circuits using AC^0_d circuits, with a correlation bound that gets significantly stronger as k gets larger.

▶ **Theorem 1.3** (AC^0_d circuits cannot approximate AC^0_{d+k} circuits)**.** *Let $n, d, k \in \mathbb{N}$ with $k \geq 3$ and $dk \leq \frac{\alpha \log n}{\log \log n}$, where $\alpha > 0$ is a suitable constant. There is an explicit AC^0_{d+k} circuit $h\colon \{0,1\}^n \to \{0,1\}$ of size $O(n)$ such that for every AC^0_d circuit $g\colon \{0,1\}^n \to \{0,1\}$, either g has size $\exp(n^{\Omega(1/d)})$, or else the following correlation bound holds:*

$$\Pr_{\mathbf{x} \in \{0,1\}^n}[g(\mathbf{x}) = h(\mathbf{x})] \leq \frac{1}{2} + \exp\left(-\frac{1}{d} \cdot \Omega(\log n)^{k-1}\right).$$

Our hard function h is the XOR of approximately $\log^{k-2} n$ many copies of Håstad, Rossman, Servedio, and Tan's hard function [39]. By combining Theorem 1.3 with the Nisan-Wigderson framework [57] and a reduction due to Li and Zuckerman [54], we obtain new constructions of *seedless randomness extractors* that are computable by small $\mathsf{AC}^0_{d+O(1)}$ circuits and that can extract from sources that are "recognizable" by large AC^0_d circuits. See the full version of this paper for details [41].

1.3.2 Correlation Bounds for XOR of Majority

Our second application of our hardness amplification technique concerns the n-bit majority function (MAJ_n). It is well known that the majority function is moderately hard for AC^0 circuits and more generally for $\mathsf{AC}^0[\oplus]$ circuits, i.e., AC^0 circuits augmented with parity gates.[4] Specifically, based on the seminal works of Razborov and Smolensky [60, 66, 67], we have the following correlation bound.

▶ **Theorem 1.4** (Majority is moderately hard for $\mathsf{AC}^0_d[\oplus]$ circuits)**.** *Let $n, d, S \in \mathbb{N}$ with $S \geq n$. Let $g\colon \{0,1\}^n \to \{0,1\}$ be an $\mathsf{AC}^0_d[\oplus]$ circuit of size S. Then*

$$\Pr_{\mathbf{x} \in \{0,1\}^n}[g(\mathbf{x}) = \mathsf{MAJ}_n(\mathbf{x})] \leq \frac{1}{2} + \frac{O(\log S)^{d-1}}{\sqrt{n}}.$$

We emphasize that we are considering the problem of computing the majority function on a $(1/2 + \varepsilon)$-fraction of n-bit inputs, which is distinct from the perhaps more famous "promise majority" problem in which we wish to compute the majority function on all inputs with relative Hamming weight outside the interval $1/2 \pm \varepsilon$. It seems that O'Donnell and Wimmer were the first to explicitly consider correlation bounds for the majority function [58].

[4] Even more generally, we can consider MOD_q gates where q is a power of a prime – but let us focus on parity gates for simplicity.

The specific quantitative bound in Theorem 1.4 is actually a log-factor improvement over what was known before, to the best of our knowledge. We therefore include a proof of Theorem 1.4 in the full version of this paper [41, Appendix A]. (We also present a matching AC^0 construction based on prior work, showing that Theorem 1.4 is tight.) That being said, our main focus is on the qualitative distinction between functions that are "moderately hard" and functions that are "very hard." The fact that the majority function is moderately hard for $\mathsf{AC}^0[\oplus]$ circuits – for example, the correlation bound above is $\widetilde{\Theta}(1/\sqrt{n})$ in the constant-depth polynomial-size regime – was already well-understood prior to this work.

Remarkably, this weak correlation bound is the best bound known on the correlation between $\mathsf{AC}^0[\oplus]$ circuits and any hard function in NP.[5] It is a major open problem to construct an explicit function that is provably "very hard" for $\mathsf{AC}^0[\oplus]$ circuits. The function $\mathsf{MAJ}_n^{\oplus t}$, perhaps with $t = \mathrm{polylog}(n)$, seems like a reasonable candidate.

Chattopadhyay, Hatami, Hosseini, Lovett, and Zuckerman recently proved that XORing amplifies the hardness of MAJ_n for constant-degree $\mathbb{F}_2$-polynomials [18], which can be considered a special case of polynomial-size $\mathsf{AC}^0_2[\oplus]$ circuits. In this work, we consider a different special case of $\mathsf{AC}^0[\oplus]$ circuits, namely AC^0 circuits. Our contribution is a proof that XORing amplifies the hardness of MAJ_n for AC^0 circuits.

▶ **Theorem 1.5** ($\mathsf{MAJ}_n^{\oplus t}$ is hard for AC^0_d circuits)**.** *Let $n, t, d, S \in \mathbb{N}$ and let $g\colon \{0,1\}^{nt} \to \{0,1\}$ be an AC^0_d circuit of size S. Then*

$$\Pr_{\mathbf{x}\in\{0,1\}^{nt}}\left[g(\mathbf{x}) = \mathsf{MAJ}_n^{\oplus t}(\mathbf{x})\right] \le \frac{1}{2} + \left(\frac{O(\log S)^{d-1}}{\sqrt{n}}\right)^t.$$

1.4 Our Technique

1.4.1 XOR Lemmas for Decision Trees

Our correlation bounds are based on XOR lemmas for *decision trees*. Before explaining the connection between AC^0 circuits and decision trees, let us discuss the XOR lemmas for decision trees themselves – a fascinating subject in its own right. Let h be a Boolean function that is moderately hard for shallow decision trees: every depth-D decision tree agrees with h on at most a $(1/2 + \varepsilon)$-fraction of inputs.

It is not hard to show that decision trees *of that same depth D* can compute $h^{\oplus t}$ on at most a $(1/2 + \varepsilon')$-fraction of inputs, where $\varepsilon' = \frac{1}{2} \cdot (2\varepsilon)^t$. (For example, this is a special case of Shaltiel's analysis of "fair" decision trees [61].) It turns out that a slight generalization of that simple analysis suffices for proving our correlation bound for depth reduction within AC^0 (Theorem 1.3).

On the other hand, to get the best parameters in Theorem 1.5 (on the hardness of $\mathsf{MAJ}_n^{\oplus t}$), it turns out that we need a more sophisticated XOR lemma for decision trees, in which we allow the tree attempting to compute $h^{\oplus t}$ to have depth significantly larger than D.

This problem has been previously studied by Drucker [26]. Focusing on one setting of parameters, Drucker showed that for every constant $\alpha > 0$, there is a value $D' = \Omega(Dt)$ such that trees of depth D' cannot compute $h^{\oplus t}$ on more than a $(1/2 + \varepsilon')$-fraction of inputs, where $\varepsilon' = O(\varepsilon)^{(1-\alpha)\cdot t}$ [26]. Although it comes close, this result is not quite sufficient to prove Theorem 1.5 because of the $(1-\alpha)$-factor loss in the exponent. Furthermore, unfortunately, the $(1-\alpha)$-factor loss is unavoidable in general, due to counterexamples

[5] If we permit hard functions that satisfy less stringent explicitness conditions, then better correlation bounds are known against $\mathsf{AC}^0[\oplus]$ and even stronger classes [77, 23, 22, 19].

identified by Shaltiel [61]. The idea behind these counterexamples is that although h is hard for decision trees of depth D, it might nevertheless be easy for decision trees of depth $D+1$. In this case, for any constant $c > 0$, a decision tree of depth cDt can successfully compute h on $\Omega(t)$ independent instances.

To circumvent Shaltiel's counterexamples [61], we strengthen the assumption. We assume that h is moderately hard for depth-D decision trees *for all D simultaneously*, with a correlation bound ε that scales with the depth D according to some log-concave function $\varepsilon(D)$. Under this assumption, we prove the decision trees of depth $\Omega(Dt)$ have correlation at most $O(\varepsilon)^t$ with $h^{\oplus t}$.

▶ **Lemma 1.6** (XOR lemma for decision trees under a robust hardness assumption). *Let $h\colon \{0,1\}^n \to \{0,1\}$ be a function and let $\varepsilon\colon [0,\infty) \to (0,\infty)$ be a log-concave function. Assume that for every $D \in \mathbb{N}$ and every decision tree $T\colon \{0,1\}^n \to \{0,1\}$ of depth at most D, we have*

$$\Pr_{\mathbf{x}\in\{0,1\}^n}[T(\mathbf{x}) = h(\mathbf{x})] \le \frac{1}{2} + \varepsilon(D).$$

Then for every $D, t \in \mathbb{N}$ and every decision tree $T\colon \{0,1\}^{nt} \to \{0,1\}$ of depth at most $Dt/2$, we have

$$\Pr_{\mathbf{x}\in\{0,1\}^{nt}}[T(\mathbf{x}) = h^{\oplus t}(\mathbf{x})] \le \frac{1}{2} + O(\varepsilon(D))^t.$$

(See Lemma 3.2 for a more general statement.)

1.4.2 Amplifying the Average-Case Depth Hierarchy Theorem

Now we briefly explain how we use an XOR lemma for decision trees to prove Theorem 1.3 (our correlation bound for depth reduction within AC^0). Our analysis builds on Håstad, Rossman, Servedio, and Tan's proof of the average-case depth hierarchy theorem [39]. Recall that their lower bound proof is based on the concept of *random projections*, which generalize traditional random restrictions. (A traditional *restriction* assigns values to some input variables while keeping others "alive." A *projection* can additionally merge living variables.) To prove that their hard function h is moderately hard for AC^0_d circuits, Håstad, Rossman, Servedio, and Tan carefully designed a distribution $\mathcal{R}$ over projections and a distribution μ over inputs and showed the following [39].

1. (Completion to the uniform distribution.) For every function $f\colon \{0,1\}^n \to \{0,1\}$, plugging a uniform random $\mathbf{x} \in \{0,1\}^n$ into f is equivalent to first sampling a projection $\boldsymbol{\pi} \sim \mathcal{R}$, then independently sampling an input $\mathbf{y} \sim \mu$, and finally plugging $\mathbf{y}$ into $f|_{\boldsymbol{\pi}}$.
2. (Simplification.) For every AC^0_d circuit g, either g has size $\exp(n^{\Omega(1/d)})$, or else with high probability over $\boldsymbol{\pi} \sim \mathcal{R}$, the circuit g *simplifies* under $\boldsymbol{\pi}$ in the sense that $g|_{\boldsymbol{\pi}}$ can be computed by a shallow decision tree.
3. (Maintaining structure.) With high probability over $\boldsymbol{\pi} \sim \mathcal{R}$, the hard function h *maintains structure* in the sense that $h|_{\boldsymbol{\pi}}$ is moderately hard for shallow decision trees with respect to μ.

Taken together, the three steps above imply that h is moderately hard for AC^0_d circuits with respect to a uniform random input. We call this proof structure the *random simplification method* for proving correlation bounds.

As mentioned previously, our hard function is $h^{\oplus t}$, where h is Håstad, Rossman, Servedio, and Tan's hard function and $t \approx \log^{k-2} n$. To prove that $h^{\oplus t}$ is very hard for AC^0_d circuits, we use the random simplification method. We apply $\mathcal{R}$ to each of the t input blocks of $h^{\oplus t}$ independently. By Håstad, Rossman, Servedio, and Tan's analysis [39], each copy of h is likely to be moderately hard for shallow decision trees after the projection. Therefore, by a suitable XOR lemma for decision trees, $h^{\oplus t}$ is likely to be *very* hard for shallow decision trees after the projection. Meanwhile, Håstad, Rossman, Servedio, and Tan's simplification arguments [39] extend to the case of several independent copies of $\mathcal{R}$, completing the proof.

1.4.3 Amplifying the Hardness of the Majority Function

There are at least three known proofs that the majority function is moderately hard for AC^0 circuits: one using the Razborov-Smolensky method [27, 50, 41], one due to O'Donnell and Wimmer [58], and one due to Tal [69]. However, none of these proofs fits into our framework of "random simplification arguments," so it is not clear how to combine them with our amplification technique. (The latter two proofs do use switching lemmas, but only in an indirect Fourier-analytic way.) For this reason, in the full version of this paper [41, §5.1], we present yet another proof that the majority function is moderately hard for AC^0_d circuits. Our new proof does fit into our "random simplification argument" framework, and furthermore, the "robust hardness assumption" of Lemma 1.6 is satisfied in our proof. These features of our proof enable us to apply our new XOR lemma for decision trees to complete our analysis of $\mathsf{MAJ}_n^{\oplus t}$.

1.5 Related Work

Goldwasser, Gutfreund, Healy, Kaufman, and Rothblum designed a method for converting worst-case hardness into moderate average-case hardness in the context of weak circuit classes [30], which complements our work in some ways. One contrast between their work and ours is that they merely construct a hard function with a very weak explicitness guarantee, namely membership in EXP, whereas we study an extremely explicit hardness amplification method, namely XORing. More recently, Chen, Lu, Lyu, and Oliveira developed a method for constructing very hard functions for weak circuit classes starting from relatively weak assumptions [20] – but once again, their hard functions only satisfy weak explicitness guarantees such as membership in E.

A long sequence of works has established strong bounds on the correlation between the parity function and AC^0 circuits [28, 1, 80, 35, 36, 7, 49, 75, 9, 44, 38]. One of these works, by Klivans [49], is especially relevant for us. Klivans' proof is based on a result by Aspnes, Beigel, Furst, and Rudich, who showed that if g is a $\mathsf{MAJ} \circ \mathsf{AC}^0_d$ circuit, then either g has size $\exp(n^{\Omega(1/d)})$, or else g disagrees with the parity function on a constant fraction of inputs [6]. Klivans combined this result with Yao's XOR Lemma to re-prove a strong (albeit not optimal) bound on the correlation between AC^0_d circuits and the parity function [49]. Klivans' proof is the only prior work we are aware of that uses hardness amplification methods to prove an unconditional AC^0 circuit lower bound.

Many prior works have studied XOR lemmas for various types of decision trees, along with the closely related "direct product" and "direct sum" problems [45, 12, 56, 61, 48, 68, 5, 46, 26, 64, 52, 13, 14, 17]. However, as far as we are aware, we are the first to consider the case that we have hardness for all depths simultaneously.

1.6 Organization

After some preliminaries, we present our XOR lemma for decision trees (Lemma 1.6) in Section 3. Then, in Section 4, we present general lemmas showing that XORing amplifies hardness whenever the hardness is proved via the random simplification method. The proofs of our main results (Theorem 1.3 and Theorem 1.5) are omitted from this extended abstract, but they can be found in the full version of this paper [41].

2 Preliminaries

We write $\mathbb{N}$ to denote the set of non-negative integers.

2.1 Boolean Functions

In the introduction, we worked with functions $f\colon \{0,1\}^n \to \{0,1\}$. Going forward, it will be more convenient to encode a bit $b \in \{0,1\}$ as the value $(-1)^b$. Thus, we will work with functions $f\colon \{\pm 1\}^n \to \{\pm 1\}$. We continue to use the notation $f^{\oplus t}$, but now $f^{\oplus t}$ denotes the *product* of t copies of f on independent inputs.

We use the following notation to describe decision trees.

▶ **Definition 2.1** (Decision trees)**.** *For a function $f\colon \{\pm 1\}^n \to \{\pm 1\}$, we define* $\mathrm{DTDepth}(f)$ *to be the minimum depth of a decision tree computing f. In the other direction, for a parameter $D \in \mathbb{N}$, we define* $\mathrm{DTDepth}[D]$ *to be the class of all functions $f\colon \{\pm 1\}^n \to \{\pm 1\}$ that can be computed by depth-D decision trees. (The parameter n will always be clear from context.)*

2.2 Probability and Correlation

We denote random variables using boldface. We write $\mathbf{x} \sim \mu$ to indicate that the random variable $\mathbf{x}$ is sampled from the distribution μ. If $\mu, \widetilde{\mu}$ are discrete probability distributions over some set Ω, then we consider the "total variation distance" between μ and $\widetilde{\mu}$ to be

$$\max_{S \subseteq \Omega}(|\Pr[\mathbf{x} \in S] - \Pr[\widetilde{\mathbf{x}} \in S]|),$$

where $\mathbf{x} \sim \mu$ and $\widetilde{\mathbf{x}} \sim \widetilde{\mu}$. We also rely on the following alternative notion of "distance" between probability distributions.

▶ **Definition 2.2** (Max-divergence)**.** *Let μ and $\widetilde{\mu}$ be discrete probability distributions over some set Ω. The* max-divergence *of $\widetilde{\mu}$ from μ is defined by*

$$D_\infty(\widetilde{\mu} \parallel \mu) = \ln\left(\max_{x \in \Omega}\left(\frac{\Pr[\widetilde{\mathbf{x}} = x]}{\Pr[\mathbf{x} = x]}\right)\right),$$

where $\mathbf{x} \sim \mu$ and $\widetilde{\mathbf{x}} \sim \widetilde{\mu}$.

Max-divergence and total variation distance are related by the following lemma.

▶ **Lemma 2.3** (Low max-divergence ⇒ low total variation distance)**.** *Let μ and $\widetilde{\mu}$ be discrete probability distributions over the same set Ω. Let $\varepsilon = D_\infty(\widetilde{\mu} \parallel \mu)$. There exists a probability distribution μ' such that μ can be written as a convex combination $\mu = (1-\varepsilon)\cdot\widetilde{\mu} + \varepsilon \cdot \mu'$. Moreover, the total variation distance between μ and $\widetilde{\mu}$ is at most ε.*

Proof. If $\varepsilon = 0$, the lemma is trivial, so assume $\varepsilon > 0$. For each $x \in \Omega$, define

$$p(x) = \frac{\Pr[\mathbf{x} = x] - (1-\varepsilon)\Pr[\widetilde{\mathbf{x}} = x]}{\varepsilon},$$

where $\mathbf{x} \sim \mu$ and $\widetilde{\mathbf{x}} \sim \widetilde{\mu}$. Then $\sum_{x \in \Omega} p(x) = 1$. Furthermore, $p(x) \geq 0$, because

$$(1-\varepsilon)\Pr[\widetilde{\mathbf{x}} = x] \leq (1-\varepsilon)\cdot e^{\varepsilon}\cdot \Pr[\mathbf{x} = x] \leq \Pr[\mathbf{x} = x].$$

Therefore, $p(\cdot)$ is a probability mass function, and we can let μ' be the corresponding probability distribution. For the "moreover" part, observe that for any $S \subseteq \Omega$, we have

$$\Pr[\mathbf{x} \in S] = (1 - \varepsilon) \cdot \Pr[\widetilde{\mathbf{x}} \in S] + \varepsilon \cdot \Pr[\mathbf{x}' \in S],$$

where $\mathbf{x}' \sim \mu'$. Therefore,

$$\Pr[\mathbf{x} \in S] \leq \Pr[\widetilde{\mathbf{x}} \in S] + \varepsilon \cdot \Pr[\mathbf{x}' \in S] \leq \Pr[\widetilde{\mathbf{x}} \in S] + \varepsilon,$$

and

$$\Pr[\mathbf{x} \in S] \geq (1 - \varepsilon) \cdot \Pr[\widetilde{\mathbf{x}} \in S] \geq \Pr[\widetilde{\mathbf{x}} \in S] - \varepsilon. \qquad ◀$$

We use the following notation for product distributions.

▶ **Definition 2.4** (Tensor product of probability distributions). *Let $\mu_1, \dots, \mu_t$ be probability distributions over the spaces $\Omega_1, \dots, \Omega_t$. Sample $\mathbf{x}_1 \sim \mu_1, \dots, \mathbf{x}_t \sim \mu_t$ independently. The* tensor product $\mu_1 \otimes \cdots \otimes \mu_t$ *is the probability distribution of $(\mathbf{x}_1, \dots, \mathbf{x}_t)$. As a special case, we define*

$$\mu^{\otimes t} = \underbrace{\mu \otimes \mu \otimes \cdots \otimes \mu}_{t \text{ copies}}.$$

We use the following standard definition to reason about average-case hardness of $\{\pm 1\}$-valued functions.

▶ **Definition 2.5** (Correlation). *Let $g, h \colon \{\pm 1\}^n \to \mathbb{R}$ be functions and let μ be a distribution over $\{\pm 1\}^n$. We define*

$$\mathsf{Corr}_\mu(g, h) = \mathop{\mathbb{E}}_{\mathbf{x} \sim \mu}[g(\mathbf{x}) \cdot h(\mathbf{x})].$$

More generally, if $\mathcal{C}$ is a class of functions $g \colon \{\pm 1\}^n \to \mathbb{R}$, then we define

$$\mathsf{Corr}_\mu(\mathcal{C}, h) = \max_{g \in \mathcal{C}} \mathsf{Corr}_\mu(g, h).$$

If μ is omitted, then by default it is assumed to be the uniform distribution over $\{\pm 1\}^n$.

If g and h are $\{\pm 1\}$-valued, then a bound $|\mathsf{Corr}(g, h)| \leq \varepsilon$ is equivalent to the statement that g agrees with h on at most a $(1/2 + \varepsilon/2)$-fraction of inputs, because for any two $\{0, 1\}$-valued random variables $\mathbf{a}, \mathbf{b}$, we have $\Pr[\mathbf{a} = \mathbf{b}] = \frac{1}{2} + \frac{1}{2}\mathbb{E}[(-1)^{\mathbf{a}} \cdot (-1)^{\mathbf{b}}]$.

2.3 Generalized Restrictions

To formulate our hardness amplification technique in the clearest and most general way possible, we work with a notion of *generalized restrictions* that includes restrictions and projections as special cases. A generalized restriction, formally defined below, consists of an arbitrary "preprocessing" step that can be applied to a Boolean function of interest.

▶ **Definition 2.6** (Generalized restriction). *A* generalized restriction *is a function $\pi \colon \{\pm 1\}^r \to \{\pm 1\}^n$. If $f \colon \{\pm 1\}^n \to \{\pm 1\}$ is a Boolean function, then we define $g|_\pi$ to be the composition $g \circ \pi$. That is, $g|_\pi \colon \{\pm 1\}^r \to \{\pm 1\}$ is given by $g|_\pi(x) = g(\pi(x))$.*

Traditional restrictions can be viewed as a special case of generalized restrictions as follows.

▶ **Definition 2.7** (Traditional restrictions as generalized restrictions). *A* restriction *is a string $\rho \in \{+1, -1, \star\}^n$. We identify ρ with a generalized restriction $\pi\colon \{\pm 1\}^r \to \{\pm 1\}^n$, where $r = |\rho^{-1}(\star)|$, as follows. Given $y \in \{\pm 1\}^r$, we let $\pi(y)$ be ρ, except that the i-th star is replaced with y_i for every $i \in [r]$.*

Next, we consider *distributions* over generalized restrictions, and we explain how to interpret the tensor product of such distributions.

▶ **Definition 2.8** (Tensor product of generalized restriction distributions). *Let $r, n \in \mathbb{N}$, and let $\mathcal{R}$ be a distribution over generalized restrictions $\boldsymbol{\pi}\colon \{\pm 1\}^r \to \{\pm 1\}^n$. Let $\boldsymbol{\pi}_1, \dots, \boldsymbol{\pi}_t$ be independent samples from $\mathcal{R}$, and define $\vec{\boldsymbol{\pi}}\colon \{\pm 1\}^{rt} \to \{\pm 1\}^{nt}$ by concatenating, i.e.,*

$$\vec{\boldsymbol{\pi}}(y^{(1)}, \dots, y^{(t)}) = (\boldsymbol{\pi}_1(y^{(1)}), \dots, \boldsymbol{\pi}_t(y^{(t)})).$$

Then the tensor product *$\mathcal{R}^{\otimes t}$ is the distribution of the random variable $\vec{\boldsymbol{\pi}}$.*

2.4 Logarithmic Concavity

We recall the following standard definition.

▶ **Definition 2.9** (Log-concave). *A function $f\colon [0, \infty) \to (0, \infty)$ is* log-concave *if $\log f$ is concave, i.e., for every $x, y \in [0, \infty)$ and $\lambda \in (0, 1)$, we have $f(x)^{\lambda} \cdot f(y)^{1-\lambda} \le f(\lambda x + (1-\lambda)y)$.*

If f is log-concave, then by induction on t, we have $\prod_{i=1}^t f(x_i) \le f(\bar{x})^t$ where $\bar{x} = \frac{1}{t}\sum_{i=1}^t x_i$.

3 XOR Lemmas for Decision Trees

In this section, we present our XOR lemma for decision trees. We begin by stating a simple XOR lemma, in which the decision tree attempting to compute $h^{\oplus t}$ has the same depth as the decision tree attempting to compute h.

▶ **Lemma 3.1** (Basic XOR lemma for decision trees). *Let $h_1, \dots, h_t\colon \{\pm 1\}^r \to \{\pm 1\}$ be functions, and define $h(y^{(1)}, \dots, y^{(t)}) = \prod_{i=1}^t h_i(y^{(i)})$. Let μ be a distribution over $\{\pm 1\}^r$. For every $D \in \mathbb{N}$, we have*

$$\mathsf{Corr}_{\mu^{\otimes t}}(h, \mathrm{DTDepth}[D]) \le \prod_{i=1}^t \mathsf{Corr}_{\mu}(h_i, \mathrm{DTDepth}[D]).$$

We were unable to find a reference for the specific statement of Lemma 3.1, but it has no significant novelty. It is closely related to Shaltiel's analysis of "fair" decision trees [61]. It can also be viewed as a special case of Claim 3.6 that we prove below. As discussed in Subsection 1.4, Lemma 3.1 is sufficient for our analysis of depth-d approximators to AC^0_{d+k} circuits (Theorem 1.3). However, for our analysis of $\mathsf{MAJ}_n^{\oplus t}$ (Theorem 1.5), we need a more sophisticated XOR lemma, stated next.

▶ **Lemma 3.2** (XOR lemma for decision trees under robust hardness assumptions, general version). *Let $h_1, \dots, h_t\colon \{\pm 1\}^r \to \{\pm 1\}$ be functions, and define $h(y^{(1)}, \dots, y^{(t)}) = \prod_{i=1}^t h_i(y^{(i)})$. Let $\mu_1, \dots, \mu_t$ be distributions over $\{\pm 1\}^r$, and define $\mu = \mu_1 \otimes \cdots \otimes \mu_t$. Let $\varepsilon\colon [0, \infty) \to (0, \infty)$ be a log-concave function, and assume that for every $i \in [t]$ and every $D \in \mathbb{N}$, we have*

$$\mathsf{Corr}_{\mu_i}(h_i, \mathrm{DTDepth}[D]) \le \varepsilon(D).$$

Then for every $D \in \mathbb{N}$, we have

$$\mathsf{Corr}_{\mu}(h, \mathrm{DTDepth}[Dt/2]) \le O(\varepsilon(D))^t.$$

The first step of the proof of Lemma 3.2 is the following claim, which enables us to relate the success probability of a tree to the success probabilities of its subtrees.

▷ Claim 3.3 (Law of total correlation). Let $h, T, E\colon \{\pm 1\}^r \to \{\pm 1\}$. Let μ be a distribution over $\{0,1\}^r$. For each $b \in \{\pm 1\}$, let $p_b = \Pr_{\mathbf{y}\sim\mu}[E(\mathbf{y}) = b]$, and let μ^b be the conditional distribution $(\mathbf{y} \sim \mu \mid E(\mathbf{y}) = b)$. Suppose that T can be decomposed in the form

$$T(y) = \begin{cases} T_{+1}(y) & \text{if } E(y) = +1 \\ T_{-1}(y) & \text{if } E(y) = -1 \end{cases}$$

for some $T_{+1}, T_{-1}\colon \{\pm 1\}^r \to \{\pm 1\}$. Then

$$\mathsf{Corr}_\mu(h, T) = \sum_{b \in \{\pm 1\}} p_b \cdot \mathsf{Corr}_{\mu^b}(h, T_b).$$

Proof.

$$\begin{aligned} \mathsf{Corr}_\mu(h, T) &= \mathop{\mathbb{E}}_{\mathbf{y}\sim\mu}[h(\mathbf{y}) \cdot T(\mathbf{y})] \\ &= \sum_{b \in \{\pm 1\}} p_b \cdot \mathop{\mathbb{E}}_{\mathbf{y}\sim\mu}[h(\mathbf{y}) \cdot T(\mathbf{y}) \mid E(\mathbf{y}) = b] \qquad \text{(Law of total expectation)} \\ &= \sum_{b \in \{\pm 1\}} p_b \mathop{\mathbb{E}}_{\mathbf{y}\sim\mu^b}[h(\mathbf{y}) \cdot T_b(\mathbf{y})]. \end{aligned}$$ ◁

Next, we consider the following notion of "fair" decision trees due to Shaltiel [61].

▶ **Definition 3.4** (($D_1, \dots, D_t$)-fair decision trees [61]). *Let $T\colon \{\pm 1\}^{rt} \to \{\pm 1\}$ be a decision tree and let $D_1, \dots, D_t \in \mathbb{N}$. We say that T is* $(D_1, \dots, D_t)$-fair *if for every input $\vec{y} = (y^{(1)}, \dots, y^{(t)}) \in (\{\pm 1\}^r)^t$, for every $i \in [t]$, the computation $T(\vec{y})$ makes at most D_i queries to $y^{(i)}$.*

The key to proving Lemma 3.2 is to generalize Definition 3.4 to the case of a *set* of tuples $(D_1, \dots, D_t)$.

▶ **Definition 3.5** (Q-fair decision trees). *Let $T\colon \{\pm 1\}^{rt} \to \{\pm 1\}$ be a decision tree and let $Q \subseteq \mathbb{N}^t$. We say that T is* Q-fair *if for every input $\vec{y} = (y^{(1)}, \dots, y^{(t)}) \in (\{\pm 1\}^r)^t$, there is some tuple $(D_1, \dots, D_t) \in Q$ such that for every $i \in [t]$, the computation $T(\vec{y})$ makes at most D_i queries to $y^{(i)}$.*

We emphasize that the tuple $(D_1, \dots, D_t)$ is permitted to vary from one input $\vec{y}$ to another. Therefore, the fact that a tree is Q-fair does not necessarily imply that there is some $(D_1, \dots, D_t) \in Q$ such that the tree is $(D_1, \dots, D_t)$-fair. Given the concept of Q-fairness, it is relatively straightforward to prove the following claim by induction on the depth of T. The claim generalizes the analysis by Shaltiel [61], who considered the case of $(D_1, \dots, D_t)$-fair decision trees and focused on the uniform distribution.

▷ Claim 3.6 (XOR lemma for Q-fair decision trees). Let $h_1, \dots, h_t\colon \{\pm 1\}^r \to \{\pm 1\}$ be functions, and define $h(y^{(1)}, \dots, y^{(t)}) = \prod_{i=1}^t h_i(y^{(i)})$. Let $\mu_1, \dots, \mu_t$ be distributions over $\{\pm 1\}^r$, and define $\mu = \mu_1 \otimes \cdots \otimes \mu_t$. Let $Q \subseteq \mathbb{N}^t$ and let $T\colon \{\pm 1\}^{rt} \to \{\pm 1\}$ be a Q-fair decision tree. Then

$$\mathsf{Corr}_\mu(h, T) \le \sum_{(D_1, \dots, D_t) \in Q} \prod_{i=1}^t \mathsf{Corr}_{\mu_i}(h_i, \mathrm{DTDepth}[D_i]).$$

Proof. Assume without loss of generality that T never queries the same variable twice. For the base case, if T has depth 0, then T is a constant function, so

$$|\mathsf{Corr}_\mu(h,T)| = \prod_{i=1}^t \left| \mathop{\mathbb{E}}_{\mathbf{y}^{(i)}\sim\mu_i}[h_i(y^{(i)})] \right| = \prod_{i=1}^t \mathsf{Corr}_{\mu_i}(h_i, \mathrm{DTDepth}[0]).$$

Since T is Q-fair, Q must be nonempty. The lemma follows because $\mathsf{Corr}_{\mu_i}(h_i,\mathrm{DTDepth}[0]) \leq \mathsf{Corr}_{\mu_i}(h_i,\mathrm{DTDepth}[D_i])$ for every $D_i \in \mathbb{N}$. For the inductive step, let $y_{j_*}^{(i_*)}$ be the variable queried by the root of the tree. Let T_{+1} and T_{-1} be the children of the root, corresponding to the cases $y_{j_*}^{(i_*)} = +1$ and $y_{j_*}^{(i_*)} = -1$ respectively. Define

$$Q' = \{(D_1,\dots,D_{i_*-1}, D_{i_*}-1, D_{i_*+1},\dots,D_t) : (D_1,\dots,D_t) \in Q \text{ and } D_{i_*} \neq 0\}.$$

Then T_{+1} and T_{-1} are both Q'-fair.

For each $b \in \{\pm 1\}$, define

$$p_b = \Pr_{\mathbf{y}^{(i_*)}\sim\mu_{i_*}}\left[\mathbf{y}_{j_*}^{(i_*)} = b\right].$$

Let $\mu_{i_*}^b$ be the conditional distribution $(\mathbf{y}^{(i_*)} \sim \mu_{i_*} \mid \mathbf{y}_{j_*}^{(i_*)} = b)$, and for $i \neq i_*$, let $\mu_i^b = \mu_i$. Let $\mu^b = \mu_1^b \otimes \cdots \otimes \mu_t^b$. By Claim 3.3 and the induction hypothesis, we have

$$\begin{aligned}
\mathsf{Corr}_\mu(h,T) &= \sum_{b\in\{\pm1\}} p_b \cdot \mathsf{Corr}_{\mu^b}(h,T_b) \\
&\leq \sum_{b\in\{\pm1\}} p_b \cdot \sum_{(D_1,\dots,D_t)\in Q'} \prod_{i=1}^t \mathsf{Corr}_{\mu_i^b}(h_i,\mathrm{DTDepth}[D_i]) \\
&= \sum_{(D_1,\dots,D_t)\in Q'} \left(\sum_{b\in\{\pm1\}} p_b \cdot \mathsf{Corr}_{\mu_{i_*}^b}(h_{i_*},\mathrm{DTDepth}[D_{i_*}]) \right) \cdot \Pi_{\neq i_*}(D_1,\dots,D_t)
\end{aligned}$$

where $\Pi_{\neq i_*}(D_1,\dots,D_t) = \prod_{i\in[t], i\neq i_*} \mathsf{Corr}_{\mu_i}(h_i,\mathrm{DTDepth}[D_i])$. Now we bound the inner sum. By Claim 3.3, for any D_{i_*}, we have

$$\mathsf{Corr}_{\mu_{i_*}}(h_{i_*},\mathrm{DTDepth}[D_{i_*}+1]) \geq \sum_{b\in\{\pm1\}} p_b \cdot \mathsf{Corr}_{\mu_{i_*}^b}(h_{i_*},\mathrm{DTDepth}[D_{i_*}]),$$

because we can approximate h_{i_*} with respect to μ_{i_*} by first querying $y_{j_*}^{(i_*)}$ and then using optimal subtrees of depth D_{i_*}. For every $(D_1,\dots,D_t) \in Q'$, we have $(D_1,\dots,D_{i_*-1}, D_{i_*}+1, D_{i_*+1},\dots,D_t) \in Q$. Therefore,

$$\mathsf{Corr}_\mu(h,T) \leq \sum_{(D_1,\dots,D_t)\in Q} \prod_{i=1}^t \mathsf{Corr}_{\mu_i}(h_i,\mathrm{DTDepth}[D_i]). \qquad \lhd$$

Given Claim 3.6, our XOR lemma for decision trees under a robust hardness assumption (Lemma 3.2) readily follows, as we now show.

Proof of Lemma 3.2. Let $T\colon \{\pm1\}^{rt} \to \{\pm1\}$ be a decision tree of depth at most $Dt/2$. Let Q be the set of t-tuples $(D_1,\dots,D_t) \in \mathbb{N}^t$ such that (1) $D_1 + \cdots + D_t \leq Dt$ and (2) D_i is an integer multiple of $\lceil D/2 \rceil$ for every i. We claim that T is Q-fair. Indeed, let $\vec{y} = (y^{(1)},\dots,y^{(t)})$ be any input, and let D_i be the number of queries that $T(\vec{y})$ makes to $y^{(i)}$. Let D_i' be the smallest integer multiple of $\lceil D/2 \rceil$ such that $D_i \leq D_i'$. Then $D_i' \leq D_i + (\lceil D/2 \rceil - 1)$, and hence $D_1' + \cdots + D_t' \leq Dt/2 + t \cdot (\lceil D/2 \rceil - 1) \leq Dt$, showing that $(D_1',\dots,D_t') \in Q$.

Therefore, by Claim 3.6,

$$\mathsf{Corr}_{\mu}(h, T) \leq \sum_{(D_1,\dots,D_t)\in Q} \prod_{i=1}^{t} \mathsf{Corr}_{\mu_i}(h_i, \mathrm{DTDepth}[D_i]).$$

For any $(D_1, \dots, D_t) \in Q$, we can define $(D'_1, \dots, D'_t)$ such that $D'_i \geq D_i$ and $D'_1 + \dots + D'_t$ is *exactly* Dt rather than being at most Dt. Then $\mathsf{Corr}_{\mu_i}(h_i, \mathrm{DTDepth}[D_i]) \leq \mathsf{Corr}_{\mu_i}(h_i, \mathrm{DTDepth}[D'_i])$, so

$$\begin{aligned}\mathsf{Corr}_{\mu}(h, T) &\leq \sum_{(D_1,\dots,D_t)\in Q} \prod_{i=1}^{t} \mathsf{Corr}_{\mu_i}(h_i, \mathrm{DTDepth}[D'_i]) \\ &\leq \sum_{(D_1,\dots,D_t)\in Q} \prod_{i=1}^{t} \varepsilon(D'_i) \\ &\leq \sum_{(D_1,\dots,D_t)\in Q} \varepsilon(D)^t && \text{(Log-concavity)} \\ &= |Q| \cdot \varepsilon(D)^t.\end{aligned}$$

To bound $|Q|$, observe that if $(D_1, \dots, D_t) \in Q$, then we can write $D_i = c_i \cdot \lceil D/2 \rceil$ for some nonnegative integers $c_1, \dots, c_t$. Furthermore, $Dt \geq \sum_i c_i \cdot \lceil D/2 \rceil \geq (D/2) \cdot \sum_i c_i$, so $c_1 + \dots + c_t \leq 2t$. Therefore, $|Q|$ is at most the number of ways that $2t$ can be partitioned into $t + 1$ nonnegative integers, which is precisely $\binom{3t}{t}$. Thus,

$$\mathsf{Corr}_{\mu}(h, T) \leq \binom{3t}{t} \cdot \varepsilon(D)^t \leq O(\varepsilon(D))^t. \qquad \blacktriangleleft$$

4 XOR Lemmas for the Random Simplification Method

In this section, we prove two general "XOR lemmas for the random simplification method," which formalize our hardness amplification technique. The first and simpler version is as follows.

▶ **Lemma 4.1** (XOR lemma for the random simplification method, basic version). *Let $n, t, r, D \in \mathbb{N}$ and $\varepsilon, \delta > 0$. Let $h\colon \{\pm 1\}^n \to \{\pm 1\}$ and $g\colon \{\pm 1\}^{nt} \to \{\pm 1\}$ be Boolean functions, let $\mathcal{R}$ be a distribution over generalized restrictions $\boldsymbol{\pi}\colon \{\pm 1\}^r \to \{\pm 1\}^n$, let μ be a distribution over $\{\pm 1\}^r$, and assume the following.*

1. *(The distribution μ completes $\mathcal{R}$ to the uniform distribution.) If we sample $\boldsymbol{\pi} \sim \mathcal{R}$ and $\mathbf{y} \sim \mu$ independently, then $\boldsymbol{\pi}(\mathbf{y})$ is a uniform random element of $\{\pm 1\}^n$.*
2. *(The function g simplifies under $\mathcal{R}^{\otimes t}$.) We have*

$$\Pr_{\vec{\boldsymbol{\pi}} \sim \mathcal{R}^{\otimes t}} \left[\mathrm{DTDepth}(g|_{\vec{\boldsymbol{\pi}}}) > D \right] \leq \delta.$$

3. *(The function h retains structure under $\mathcal{R}$.) We have*

$$\mathop{\mathbb{E}}_{\boldsymbol{\pi} \sim \mathcal{R}} \left[\mathsf{Corr}_{\mu}(h|_{\boldsymbol{\pi}}, \mathrm{DTDepth}[D]) \right] \leq \varepsilon.$$

Then $\mathsf{Corr}(g, h^{\oplus t}) \leq \varepsilon^t + \delta$.

Proof. Sample $\vec{\boldsymbol{\pi}} = (\boldsymbol{\pi}_1, \dots, \boldsymbol{\pi}_t) \sim \mathcal{R}^{\otimes t}$ and $\vec{\mathbf{y}} \sim \mu^{\otimes t}$ independently. Let $\mathbf{T}$ be $g|_{\vec{\boldsymbol{\pi}}}$ if $\mathrm{DTDepth}(g|_{\vec{\boldsymbol{\pi}}}) \leq D$; otherwise, let $\mathbf{T}$ be the constant-zero function. Assumption 1 implies that $\vec{\boldsymbol{\pi}}(\vec{\mathbf{y}})$ is distributed uniformly over $\{\pm 1\}^{nt}$. Therefore,

$$\begin{aligned}
\mathsf{Corr}(h^{\oplus t}, g) &= \mathop{\mathbb{E}}_{\vec{\boldsymbol{\pi}}}\left[\mathsf{Corr}_{\mu^{\oplus t}}(h^{\oplus t}|_{\vec{\boldsymbol{\pi}}}, g|_{\vec{\boldsymbol{\pi}}})\right] && \text{(Assumption 1)}\\
&\leq \delta + \mathop{\mathbb{E}}_{\vec{\boldsymbol{\pi}}}\left[\mathsf{Corr}_{\mu^{\oplus t}}(h^{\oplus t}|_{\vec{\boldsymbol{\pi}}}, \mathbf{T})\right] && \text{(Assumption 2)}\\
&\leq \delta + \mathop{\mathbb{E}}_{\vec{\boldsymbol{\pi}}}\left[\mathsf{Corr}_{\mu^{\oplus t}}(h^{\oplus t}|_{\vec{\boldsymbol{\pi}}}, \mathrm{DTDepth}[D])\right] && \\
&\leq \delta + \mathop{\mathbb{E}}_{\vec{\boldsymbol{\pi}}}\left[\prod_{i=1}^{t} \mathsf{Corr}_{\mu}(h|_{\boldsymbol{\pi}_i}, \mathrm{DTDepth}[D])\right] && \text{(Lemma 3.1)}\\
&= \delta + \left(\mathop{\mathbb{E}}_{\boldsymbol{\pi} \sim \mathcal{R}}\left[\mathsf{Corr}_{\mu}(h|_{\boldsymbol{\pi}}, \mathrm{DTDepth}[D])\right]\right)^t && \text{(Independence)}\\
&\leq \delta + \varepsilon^t && \text{(Assumption 3.)}
\end{aligned}$$

◀

In the full version of this paper [41], we review the basic structure of Håstad, Rossman, Servedio, and Tan's proof of the average-case depth hierarchy theorem [39] and explain how it fits into the framework of Lemma 4.1. As a result, we are able to use Lemma 4.1 to prove our result about the average-case hardness of AC^0_{d+k} circuits for AC^0_d circuits (Theorem 1.3).

In this extended abstract, let us focus on the hardness amplification technique itself. The conclusion of Lemma 4.1 is $\mathsf{Corr}(g, h^{\oplus t}) \leq \varepsilon^t + \delta$. The "$+\ \delta$" term is unfortunate, since it does not improve with increasing t. To address this weakness, we now prove a more sophisticated version of Lemma 4.1 in which the correlation bound is $O(\varepsilon)^t$, with no "$+\ \delta$" term, albeit under stronger assumptions.

▶ **Lemma 4.2** (Tighter XOR lemma for the random simplification method)**.** *Let $n, t \in \mathbb{N}$ and let $h\colon \{\pm 1\}^n \to \{\pm 1\}$ be a Boolean function. Let $\mathcal{C}$ be a class of Boolean functions $g\colon \{\pm 1\}^{nt} \to \{\pm 1\}$ that is closed under restrictions.*[6] *Let $r \in \mathbb{N}$, let $\mathcal{R}$ be a distribution over generalized restrictions $\boldsymbol{\pi}\colon \{\pm 1\}^r \to \{\pm 1\}^n$, and let μ be a distribution over $\{\pm 1\}^r$. Let $\varepsilon > 0$, and assume the following.*

1. *(The distribution μ approximately completes $\mathcal{R}$ to the uniform distribution.) If we sample $\boldsymbol{\pi} \sim \mathcal{R}$ and $\mathbf{y} \sim \mu$ independently, and we sample $\mathbf{x} \in \{\pm 1\}^n$ uniformly at random, then $D_\infty(\boldsymbol{\pi}(\mathbf{y}) \parallel \mathbf{x}) \leq \varepsilon$.*
2. *(The class $\mathcal{C}$ simplifies under $\mathcal{R}^{\otimes t}$.) For every $g \in \mathcal{C}$ and every $D \in \mathbb{N}$, we have*
$$\Pr_{\vec{\boldsymbol{\pi}} \sim \mathcal{R}^{\otimes t}}\left[\mathrm{DTDepth}(g|_{\vec{\boldsymbol{\pi}}}) \geq D\right] \leq 2^{t-D}.$$
3. *(The function h retains structure under $\mathcal{R}$.) For every $D \in \mathbb{N}$ and every $\pi \in \mathrm{Supp}(\mathcal{R})$, we have*
$$\mathsf{Corr}_{\mu}(h|_{\pi}, \mathrm{DTDepth}[D]) \leq \varepsilon \cdot 2^{D/3}.$$

Then $\mathsf{Corr}(\mathcal{C}, h^{\oplus t}) \leq O(\varepsilon)^t$.

Proof. Fix any $g \in \mathcal{C}$. Our job is to analyze the correlation between g and $h^{\oplus t}$ under a uniform random input. By Lemma 2.3, we can sample a uniform random input by the following procedure.

[6] Every function in $\mathcal{C}$ has domain $\{\pm 1\}^{nt}$. When we say that $\mathcal{C}$ is closed under restrictions, we are thinking of a restriction of $g \in \mathcal{C}$ as another function on nt bits that ignores some of its input variables.

1. Sample $\vec{\boldsymbol{\pi}} = (\boldsymbol{\pi}_1, \dots, \boldsymbol{\pi}_t) \sim \mathcal{R}^{\otimes t}$.
2. Sample $\vec{\mathbf{y}} = (\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(t)}) \sim \mu^{\otimes t}$.
3. Sample $\vec{\mathbf{e}} = (\mathbf{e}^{(1)}, \dots, \mathbf{e}^{(t)}) \sim (\mu')^{\otimes t}$, where μ' is the distribution over $\{\pm 1\}^n$ from Lemma 2.3.
4. Sample $\mathbf{I} \subseteq [t]$ where $\Pr[i \in \mathbf{I}] = 1 - \varepsilon$ independently for every i.
5. Output the string $\vec{\mathbf{x}} = (\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(t)}) \in \{\pm 1\}^{nt}$, where

$$\mathbf{x}^{(i)} = \begin{cases} \boldsymbol{\pi}_i(\mathbf{y}^{(i)}) & \text{if } i \in \mathbf{I} \\ \mathbf{e}^{(i)} & \text{if } i \notin \mathbf{I}. \end{cases}$$

Let $\mathbf{g} \colon \{\pm 1\}^{nt} \to \{\pm 1\}$ be the function obtained from g by plugging $\mathbf{e}^{(i)}$ into each block $i \notin \mathbf{I}$ and leaving the blocks in $\mathbf{I}$ alive. Since $\mathbf{g}$ ignores the variables in blocks outside $\mathbf{I}$, we have

$$g(\vec{\mathbf{x}}) = \mathbf{g}|_{\vec{\boldsymbol{\pi}}}(\vec{\mathbf{y}}).$$

Similarly, define $\mathbf{h} \colon \{\pm 1\}^{nt} \to \{\pm 1\}$ by the formula

$$\mathbf{h}(\vec{x}) = \left(\prod_{i \in \mathbf{I}} h_i(x^{(i)})\right) \cdot \left(\prod_{i \notin \mathbf{I}} h_i(\mathbf{e}^{(i)})\right),$$

so that $h^{\oplus t}(\vec{\mathbf{x}}) = \mathbf{h}|_{\vec{\boldsymbol{\pi}}}(\vec{\mathbf{y}})$. That way,

$$\mathsf{Corr}(g, h^{\oplus t}) = \mathbb{E}[g(\vec{\mathbf{x}}) \cdot h^{\oplus t}(\vec{\mathbf{x}})] = \mathop{\mathbb{E}}_{\mathbf{I}, \vec{\mathbf{e}}, \vec{\boldsymbol{\pi}}}[\mathsf{Corr}_{\mu^{\otimes t}}(\mathbf{g}|_{\vec{\boldsymbol{\pi}}}, \mathbf{h}|_{\vec{\boldsymbol{\pi}}})].$$

Let $\mathbf{D} = \lfloor \mathrm{DTDepth}(\mathbf{g}|_{\vec{\boldsymbol{\pi}}}) / |\mathbf{I}| \rfloor$. Then

$$\mathop{\mathbb{E}}_{\mathbf{I}, \vec{\mathbf{e}}, \vec{\boldsymbol{\pi}}}[\mathsf{Corr}_{\mu^{\otimes t}}(\mathbf{g}|_{\vec{\boldsymbol{\pi}}}, \mathbf{h}|_{\vec{\boldsymbol{\pi}}})] \le \mathop{\mathbb{E}}_{\mathbf{I}, \vec{\mathbf{e}}, \vec{\boldsymbol{\pi}}}[\mathsf{Corr}_{\mu^{\otimes t}}(\mathbf{h}|_{\vec{\boldsymbol{\pi}}}, \mathrm{DTDepth}[(\mathbf{D} + 1) \cdot |\mathbf{I}|])].$$

Let $\boldsymbol{\pi}_{\mathbf{I}} = (\boldsymbol{\pi}_i)_{i \in \mathbf{I}}$, and define $h_{\mathbf{I}} \colon \{\pm 1\}^{n|\mathbf{I}|} \to \{\pm 1\}$ by $h_{\mathbf{I}}((x^{(i)})_{i \in \mathbf{I}}) = \prod_{i \in \mathbf{I}} h(x^{(i)})$. Then for any fixing of $\mathbf{I}, \vec{\mathbf{e}}, \vec{\boldsymbol{\pi}}$, we have

$$\mathsf{Corr}_{\mu^{\otimes t}}(\mathbf{h}|_{\vec{\boldsymbol{\pi}}}, \mathrm{DTDepth}[(\mathbf{D} + 1) \cdot |\mathbf{I}|]) = \mathsf{Corr}_{\mu^{\otimes |\mathbf{I}|}}(h_{\mathbf{I}}|_{\boldsymbol{\pi}_{\mathbf{I}}}, \mathrm{DTDepth}[(\mathbf{D} + 1) \cdot |\mathbf{I}|]).$$

Now we apply Lemma 3.2. For each $i \in \mathbf{I}$ and each $D \in \mathbb{N}$, we have

$$\mathsf{Corr}_{\mu}(h|_{\boldsymbol{\pi}_i}, \mathrm{DTDepth}[D]) \le \varepsilon \cdot 2^{D/3}.$$

Furthermore, the function $\varepsilon(D) = \varepsilon \cdot 2^{D/3}$ is log-concave. Therefore, Lemma 3.2 guarantees that

$$\mathsf{Corr}_{\mu^{\otimes |\mathbf{I}|}}(h_{\mathbf{I}}|_{\boldsymbol{\pi}_{\mathbf{I}}}, \mathrm{DTDepth}[(\mathbf{D} + 1) \cdot |\mathbf{I}|]) \le O(\varepsilon \cdot 2^{2 \cdot (\mathbf{D} + 1)/3})^{|\mathbf{I}|} = O(\varepsilon \cdot 2^{2\mathbf{D}/3})^{|\mathbf{I}|}.$$

Thus, overall, we get

$$\mathsf{Corr}(g, h^{\oplus t}) \le \mathop{\mathbb{E}}_{\mathbf{I}, \vec{\mathbf{e}}, \vec{\boldsymbol{\pi}}}\left[O\left(\varepsilon \cdot 2^{2\mathbf{D}/3}\right)^{|\mathbf{I}|}\right].$$

Now consider any fixing of $\mathbf{I}$ and $\vec{\mathbf{e}}$. Since $\mathcal{C}$ is closed under restrictions, $\mathbf{g} \in \mathcal{C}$. Therefore, our simplification assumption tells us that for every $D \in \mathbb{N}$, we have

$$\Pr_{\vec{\boldsymbol{\pi}}}[\mathbf{D} = D] \le 2^{t - D \cdot |\mathbf{I}|}.$$

Consequently,

$$\begin{aligned}\mathop{\mathbb{E}}_{\vec{\boldsymbol{\pi}}}\left[O\left(\varepsilon\cdot 2^{2\mathbf{D}/3}\right)^{|\mathbf{I}|}\right] &= \sum_{D=0}^{\infty}\Pr_{\vec{\boldsymbol{\pi}}}[\mathbf{D}=D]\cdot O\left(\varepsilon\cdot 2^{2D/3}\right)^{|\mathbf{I}|}\\ &\le 2^t\cdot\sum_{D=0}^{\infty}O\left(\varepsilon\cdot 2^{-D/3}\right)^{|\mathbf{I}|}\\ &\le 2^t\cdot O\left(\sum_{D=0}^{\infty}\varepsilon\cdot 2^{-D/3}\right)^{|\mathbf{I}|}\\ &= 2^t\cdot O(\varepsilon)^{|\mathbf{I}|}.\end{aligned}$$

Therefore, our overall bound is given by

$$\begin{aligned}\mathsf{Corr}(g,h^{\oplus t}) \le \mathop{\mathbb{E}}_{\mathbf{I}}[2^t\cdot O(\varepsilon)^{|\mathbf{I}|}] &= 2^t\cdot\sum_{I\subseteq[t]}\Pr[\mathbf{I}=I]\cdot O(\varepsilon)^{|I|}\\ &= 2^t\cdot\sum_{I\subseteq[t]}(1-\varepsilon)^{|I|}\cdot\varepsilon^{t-|I|}\cdot O(\varepsilon)^{|I|}\\ &\le O(\varepsilon)^t.\end{aligned}$$

◀

In the full version of this paper [41], we explain how to use Lemma 4.2 to prove our correlation bound for $\mathsf{MAJ}_n^{\oplus t}$ (Theorem 1.5).

5 Directions for Further Research

The main open question related to our work is whether XORing always amplifies hardness for AC^0 circuits (cf. Theorem 1.1). We wish to also highlight the problem of proving *tight* correlation bounds for depth reduction within AC^0 (cf. Theorem 1.3). That is, what is the correlation between linear-size AC^0_{d+k} circuits and near-exponential-size AC^0_d circuits?

For simplicity, let us consider the case that d and k are both constants. As discussed previously, the extreme case $k=1$ (i.e., using AC^0_d circuits to approximate AC^0_{d+1} circuits) is resolved by Håstad, Rossman, Servedio, and Tan's work [39] to within polynomial factors; the optimal correlation bound is $n^{\Theta(1)}$. Prior work also implies near-matching upper and lower bounds in the opposite extreme case $d=1$ (i.e., using AC^0_1 circuits to approximate AC^0_{1+k} circuits). In this case, it turns out that the optimal correlation bound is $\exp\left(-\widetilde{\Theta}(\log^k n)\right)$. (The approximators are based on the Linial-Nisan-Mansour theorem [55]; see the full version of this paper [41, Appendix B] for details.)

Based on those two extreme cases, it is tempting to conjecture that for all d and k, the optimal correlation bound should be $\exp\left(-\widetilde{\Theta}(\log^k n)\right)$, but in truth it is not at all clear that this is the best guess. Arguably the most interesting case is $k=2$, i.e., the problem of using AC^0_d circuits to approximate AC^0_{d+2} circuits. On the one hand, the best method we know for constructing such an approximator is simply to use an optimal AC^0_1 approximator. On the other hand, the best correlation bound we know for this case is Håstad, Rossman, Servedio, and Tan's bound [39]. We therefore have a considerable gap between the upper and lower correlation bounds for this case, namely $n^{-\Omega(1)}$ vs. $n^{-\widetilde{O}(\log^d n)}$.

References

1 M. Ajtai. Σ_1^1-formulae on finite structures. *Ann. Pure Appl. Logic*, 24(1):1–48, 1983. doi:10.1016/0168-0072(83)90038-6.

2 Eric Allender. A note on the power of threshold circuits. In *Proceedings of the 30th Symposium on Foundations of Computer Science (FOCS)*, pages 580–584, 1989. doi:10.1109/SFCS.1989.63538.

3 Eric Allender and Vivek Gore. A uniform circuit lower bound for the permanent. *SIAM Journal on Computing*, 23(5):1026–1049, 1994. doi:10.1137/S0097539792233907.

4 Eric Allender and Ulrich Hertrampf. Depth reduction for circuits of unbounded fan-in. *Inform. and Comput.*, 112(2):217–238, 1994. doi:10.1006/inco.1994.1057.

5 Andris Ambainis, Robert Špalek, and Ronald de Wolf. A new quantum lower bound method, with applications to direct product theorems and time-space tradeoffs. *Algorithmica*, 55(3):422–461, 2009. doi:10.1007/s00453-007-9022-9.

6 James Aspnes, Richard Beigel, Merrick Furst, and Steven Rudich. The expressive power of voting polynomials. *Combinatorica*, 14(2):135–148, 1994. doi:10.1007/BF01215346.

7 László Babai. Random oracles separate PSPACE from the polynomial-time hierarchy. *Inform. Process. Lett.*, 26(1):51–53, 1987. doi:10.1016/0020-0190(87)90036-6.

8 Louay M. J. Bazzi. Polylogarithmic independence can fool DNF formulas. *SIAM J. Comput.*, 38(6):2220–2272, 2009. doi:10.1137/070691954.

9 Paul Beame, Russell Impagliazzo, and Srikanth Srinivasan. Approximating AC^0 by small height decision trees and a deterministic algorithm for $\#\mathrm{AC}^0SAT$. In *27th Conference on Computational Complexity (CCC)*, pages 117–125, 2012. doi:10.1109/CCC.2012.40.

10 Richard Beigel, Nick Reingold, and Daniel A. Spielman. The perceptron strikes back. In *Proceedings of the 6th Annual Structure in Complexity Theory Conference (SCT)*, pages 286–291, 1991. doi:10.1109/SCT.1991.160270.

11 Richard Beigel and Jun Tarui. On ACC. *Comput. Complexity*, 4(4):350–366, 1994. doi:10.1007/BF01263423.

12 Yosi Ben-Asher and Ilan Newman. Decision trees with and, or queries. In *Proceedings of the 10th Conference on Structure in Complexity Theory (SCT)*, pages 74–81, 1995. doi:10.1109/SCT.1995.514729.

13 Shalev Ben-David and Robin Kothari. Randomized query complexity of sabotaged and composed functions. *Theory Comput.*, 14:Paper No. 5, 27, 2018. doi:10.4086/toc.2018.v014a005.

14 Eric Blais and Joshua Brody. Optimal Separation and Strong Direct Sum for Randomized Query Complexity. In *Proceedings of the 34th Computational Complexity Conference (CCC)*, pages 29:1–29:17, 2019. doi:10.4230/LIPIcs.CCC.2019.29.

15 Ravi B. Boppana. The average sensitivity of bounded-depth circuits. *Information Processing Letters*, 63(5):257–261, 1997. doi:10.1016/S0020-0190(97)00131-2.

16 Mark Braverman. Polylogarithmic independence fools AC^0 circuits. *Journal of the ACM*, 57(5), 2010.

17 Joshua Brody, Jae Tak Kim, Peem Lerdputtipongporn, and Hariharan Srinivasulu. A strong XOR lemma for randomized query complexity, 2020. arXiv:2007.05580.

18 Eshan Chattopadhyay, Pooya Hatami, Kaave Hosseini, Shachar Lovett, and David Zuckerman. XOR lemmas for resilient functions against polynomials. In *Proceedings of the 52nd Symposium on Theory of Computing (STOC)*, pages 234–246, 2020. doi:10.1145/3357713.3384242.

19 Lijie Chen. New Lower Bounds and Derandomization for ACC, and a Derandomization-Centric View on the Algorithmic Method. In *14th Innovations in Theoretical Computer Science Conference (ITCS)*, pages 34:1–34:15, 2023. doi:10.4230/LIPIcs.ITCS.2023.34.

20 Lijie Chen, Zhenjian Lu, Xin Lyu, and Igor C. Oliveira. Majority vs. approximate linear sum and average-case complexity below NC^1. In *48th International Colloquium on Automata, Languages, and Programming (ICALP)*, pages 51:1–51:20, 2021. doi:10.4230/LIPIcs.ICALP.2021.51.

21 Lijie Chen and Xin Lyu. Inverse-exponential correlation bounds and extremely rigid matrices from a new derandomized XOR lemma. In *Proceedings of the 53rd Symposium on Theory of Computing (STOC)*, pages 761–771, 2021. doi:10.1145/3406325.3451132.

22 Lijie Chen, Xin Lyu, and R. Ryan Williams. Almost-everywhere circuit lower bounds from non-trivial derandomization. In *61st Symposium on Foundations of Computer Science (FOCS)*, pages 1–12, 2020. doi:10.1109/FOCS46700.2020.00009.

23 Lijie Chen and Hanlin Ren. Strong average-case circuit lower bounds from nontrivial derandomization. *SIAM Journal on Computing*, 51(3):STOC20–115–STOC20–173, 2022. doi:10.1137/20M1364886.

24 Shiteng Chen and Periklis A. Papakonstantinou. Depth reduction for composites. *SIAM J. Comput.*, 48(2):668–686, 2019. doi:10.1137/17M1129672.

25 Yeyuan Chen, Yizhi Huang, Jiatu Li, and Hanlin Ren. Range avoidance, remote point, and hard partial truth table via satisfying-pairs algorithms. In *Proceedings of the 55th Symposium on Theory of Computing (STOC)*, pages 1058–1066, 2023. doi:10.1145/3564246.3585147.

26 Andrew Drucker. Improved direct product theorems for randomized query complexity. *Comput. Complexity*, 21(2):197–244, 2012. doi:10.1007/s00037-012-0043-7.

27 Yuval Filmus. Smolensky's lower bound. Unpublished, 2010. URL: https://yuvalfilmus.cs.technion.ac.il/Manuscripts/Smolensky.pdf.

28 Merrick Furst, James B. Saxe, and Michael Sipser. Parity, circuits, and the polynomial-time hierarchy. *Math. Systems Theory*, 17(1):13–27, 1984. doi:10.1007/BF01744431.

29 Oded Goldreich, Noam Nisan, and Avi Wigderson. On Yao's XOR-lemma. In *Studies in complexity and cryptography*, volume 6650 of *Lecture Notes in Comput. Sci.*, pages 273–301. Springer, Heidelberg, 2011. doi:10.1007/978-3-642-22670-0_23.

30 Shafi Goldwasser, Dan Gutfreund, Alexander Healy, Tali Kaufman, and Guy N. Rothblum. Verifying and decoding in constant depth. In *Proceedings of the 39th Symposium on Theory of Computing (STOC)*, pages 440–449, 2007. doi:10.1145/1250790.1250855.

31 Aryeh Grinberg, Ronen Shaltiel, and Emanuele Viola. Indistinguishability by adaptive procedures with advice, and lower bounds on hardness amplification proofs. In *Proceedings of the 59th Symposium on Foundations of Computer Science (FOCS)*, pages 956–966, 2018. doi:10.1109/FOCS.2018.00094.

32 Dan Gutfreund and Guy N. Rothblum. The complexity of local list decoding. In *Proceedings of the 12th International Conference on Randomization and Computation (RANDOM)*, pages 455–468, 2008. doi:10.1007/978-3-540-85363-3_36.

33 András Hajnal, Wolfgang Maass, Pavel Pudlák, Mario Szegedy, and György Turán. Threshold circuits of bounded depth. *Journal of Computer and System Sciences*, 46(2):129–154, 1993. doi:10.1016/0022-0000(93)90001-D.

34 Prahladh Harsha and Srikanth Srinivasan. On polynomial approximations to AC^0. *Random Structures Algorithms*, 54(2):289–303, 2019. doi:10.1002/rsa.20786.

35 Johan Håstad. Almost optimal lower bounds for small depth circuits. In *Proceedings of the 18th Symposium on Theory of Computing (STOC)*, pages 6–20, 1986. doi:10.1145/12130.12132.

36 Johan Håstad. *Computational limitations for small depth circuits.* PhD thesis, Massachusetts Institute of Technology, 1986.

37 Johan Håstad. A slight sharpening of LMN. *Journal of Computer and System Sciences*, 63(3):498–508, 2001. doi:10.1006/jcss.2001.1803.

38 Johan Håstad. On the correlation of parity and small-depth circuits. *SIAM J. Comput.*, 43(5):1699–1708, 2014. doi:10.1137/120897432.

39 Johan Håstad, Benjamin Rossman, Rocco A. Servedio, and Li-Yang Tan. An average-case depth hierarchy theorem for Boolean circuits. *J. ACM*, 64(5):Art. 35, 27, 2017. doi:10.1145/3095799.

40 Pooya Hatami, William M. Hoza, Avishay Tal, and Roei Tell. Depth-d threshold circuits vs. depth-$(d+1)$ and-or trees. In *Proceedings of the 55th Symposium on Theory of Computing (STOC)*, pages 895–904, 2023. Full version: https://eccc.weizmann.ac.il/report/2022/087/. doi:10.1145/3564246.3585216.

41 William M. Hoza. A technique for hardness amplification against AC^0. https://eccc.weizmann.ac.il/report/2023/176/, 2023.

42 Xuangui Huang and Emanuele Viola. Average-case rigidity lower bounds. In *Proceedings of the 16th International Computer Science Symposium in Russia (CSR)*, pages 186–205, 2021. doi:10.1007/978-3-030-79416-3_11.

43 Russell Impagliazzo. Hard-core distributions for somewhat hard problems. In *36th Symposium on Foundations of Computer Science (FOCS)*, pages 538–545, 1995. doi:10.1109/SFCS.1995.492584.

44 Russell Impagliazzo, William Matthews, and Ramamohan Paturi. A satisfiability algorithm for AC^0. In *Proceedings of the 23rd Symposium on Discrete Algorithms (SODA)*, pages 961–972, 2012. doi:10.1137/1.9781611973099.77.

45 Russell Impagliazzo, Ran Raz, and Avi Wigderson. A direct product theorem. In *Proceedings of 9th Annual Conference on Structure in Complexity Theory (SCT)*, pages 88–96, 1994. doi:10.1109/SCT.1994.315814.

46 Rahul Jain, Hartmut Klauck, and Miklos Santha. Optimal direct sum results for deterministic and randomized decision tree complexity. *Inform. Process. Lett.*, 110(20):893–897, 2010. doi:10.1016/j.ipl.2010.07.020.

47 Jeff Kahn, Gil Kalai, and Nathan Linial. The influence of variables on boolean functions. In *Proceedings of the 29th Symposium on Foundations of Computer Science (FOCS)*, pages 68–80, 1988. doi:10.1109/SFCS.1988.21923.

48 Hartmut Klauck, Robert Špalek, and Ronald de Wolf. Quantum and classical strong direct product theorems and optimal time-space tradeoffs. *SIAM J. Comput.*, 36(5):1472–1493, 2007. doi:10.1137/05063235X.

49 Adam R. Klivans. On the derandomization of constant depth circuits. In *Proceedings of the 5th International Workshop on Randomization and Approximation Techniques in Computer Science (RANDOM)*, pages 249–260, 2001. doi:10.1007/3-540-44666-4_28.

50 Swastik Kopparty. Lecture 4: AC^0 lower bounds and pseudorandomness. Scribe notes by Jason Perry and Brian Garnett, 2013. URL: https://sites.math.rutgers.edu/~sk1233/courses/topics-S13/lec4.pdf.

51 Swastik Kopparty and Srikanth Srinivasan. Certifying polynomials for $\mathrm{AC}^0[\oplus]$ circuits, with applications to lower bounds and circuit compression. *Theory of Computing*, 14(12):1–24, 2018. doi:10.4086/toc.2018.v014a012.

52 Troy Lee and Jérémie Roland. A strong direct product theorem for quantum query complexity. *Comput. Complexity*, 22(2):429–462, 2013. doi:10.1007/s00037-013-0066-8.

53 L. A. Levin. One way functions and pseudorandom generators. *Combinatorica*, 7(4):357–363, 1987. doi:10.1007/BF02579323.

54 Fu Li and David Zuckerman. Improved extractors for recognizable and algebraic sources. In *Proceedings of the 23rd International Conference on Randomization and Computation (RANDOM)*, pages 72:1–72:22, 2019. doi:10.4230/LIPIcs.APPROX-RANDOM.2019.72.

55 Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, Fourier transform, and learnability. *Journal of the ACM*, 40(3):607–620, 1993. doi:10.1145/174130.174138.

56 Noam Nisan, Steven Rudich, and Michael Saks. Products and help bits in decision trees. *SIAM J. Comput.*, 28(3):1035–1050, 1999. doi:10.1137/S0097539795282444.

57 Noam Nisan and Avi Wigderson. Hardness vs. randomness. *J. Comput. System Sci.*, 49(2):149–167, 1994. doi:10.1016/S0022-0000(05)80043-1.

58 Ryan O'Donnell and Karl Wimmer. Approximation by DNF: examples and counterexamples. In *Proceedings of the 34th International Colloquium on Automata, Languages and Programming (ICALP)*, pages 195–206, 2007. doi:10.1007/978-3-540-73420-8_19.

59 Alexander Razborov. A simple proof of Bazzi's theorem. *ACM Transactions on Computation Theory*, 1(1), 2009. doi:10.1145/1490270.1490273.

60 Alexander A. Razborov. Lower bounds on the size of constant-depth networks over a complete basis with logical addition. *Mathematical Notes of the Academy of Science of the USSR*, 41(4):333–338, 1987. doi:10.1007/BF01137685.

61 Ronen Shaltiel. Towards proving strong direct product theorems. *Comput. Complexity*, 12(1-2):1–22, 2003. doi:10.1007/s00037-003-0175-x.

62 Ronen Shaltiel. Is it possible to improve Yao's XOR lemma using reductions that exploit the efficiency of their oracle? *Comput. Complexity*, 32(1):Paper No. 5, 47, 2023. doi:10.1007/s00037-023-00238-9.

63 Ronen Shaltiel and Emanuele Viola. Hardness amplification proofs require majority. *SIAM J. Comput.*, 39(7):3122–3154, 2010. doi:10.1137/080735096.

64 Alexander A. Sherstov. Strong direct product theorems for quantum communication and query complexity. *SIAM J. Comput.*, 41(5):1122–1165, 2012. doi:10.1137/110842661.

65 Michael Sipser. Borel sets and circuit complexity. In *Proceedings of the 15th Symposium on Theory of Computing*, pages 61–69, 1983. doi:10.1145/800061.808733.

66 Roman Smolensky. Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *Proceedings of the 19th Symposium on Theory of Computing (STOC)*, pages 77–82, 1987. doi:10.1145/28395.28404.

67 Roman Smolensky. On representations by low-degree polynomials. In *Proceedings of 34th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 130–138, 1993. doi:10.1109/SFCS.1993.366874.

68 Robert Špalek. The multiplicative quantum adversary. In *Proceedings of the 23rd Conference on Computational Complexity (CCC)*, pages 237–248, 2008. doi:10.1109/CCC.2008.9.

69 Avishay Tal. Tight bounds on the fourier spectrum of AC0. In *Proceedings of the 32nd Computational Complexity Conference (CCC)*, pages 15:1–15:31, 2017. doi:10.4230/LIPIcs.CCC.2017.15.

70 Jun Tarui. Probabilistic polynomials, AC^0 functions and the polynomial-time hierarchy. *Theoretical Computer Science*, 113(1):167–183, 1993. doi:10.1016/0304-3975(93)90214-E.

71 Roei Tell. On implications of better sub-exponential lower bounds for $\mathcal{AC}^0$. https://sites.google.com/site/roeitell/Expositions, 2020.

72 Seinosuke Toda. PP is as hard as the polynomial-time hierarchy. *SIAM J. Comput.*, 20(5):865–877, 1991. doi:10.1137/0220053.

73 Leslie G. Valiant. Graph-theoretic arguments in low-level complexity. In *Proceedings of the 6th Symposium on Mathematical Foundations of Computer Science (MFCS)*, pages 162–176, 1977. doi:10.1007/3-540-08353-7_135.

74 Emanuele Viola. *The complexity of hardness amplification and derandomization.* PhD thesis, Harvard University, 2006.

75 Emanuele Viola. On the power of small-depth computation. *Found. Trends Theor. Comput. Sci.*, 5(1):1–72, 2009. doi:10.1561/0400000033.

76 Emanuele Viola. Selected challenges in computational lower bounds. *SIGACT News*, 48(1):39–45, March 2017. doi:10.1145/3061640.3061648.

77 Emanuele Viola. New lower bounds for probabilistic degree and ac0 with parity gates. https://eccc.weizmann.ac.il/report/2020/015/, 2020.

78 Ryan Williams. Nonuniform acc circuit lower bounds. *J. ACM*, 61(1), January 2014. doi:10.1145/2559903.

79 Andrew C. Yao. Theory and application of trapdoor functions. In *Proceedings of the 23rd Symposium on Foundations of Computer Science (FOCS)*, pages 80–91, 1982. doi:10.1109/SFCS.1982.45.

80 Andrew Chi-Chih Yao. Separating the polynomial-time hierarchy by oracles. In *26th Symposium on Foundations of Computer Science (FOCS)*, pages 1–10, 1985. doi:10.1109/SFCS.1985.49.

81 Andrew Chi-Chih Yao. On ACC and threshold circuits. In *Proceedings of the 31st Symposium on Foundations of Computer Science (FOCS)*, pages 619–627, 1990. doi:10.1109/FSCS.1990.89583.