Luby–Veličković–Wigderson Revisited: Improved Correlation Bounds and Pseudorandom Generators for Depth-Two Circuits

Rocco A. Servedio[1]
Department of Computer Science, Columbia University, New York, NY, USA
rocco@cs.columbia.edu

Li-Yang Tan[2]
Department of Computer Science, Stanford University, Stanford, California, USA
liyang@cs.stanford.edu

Abstract

We study correlation bounds and pseudorandom generators for depth-two circuits that consist of a SYM-gate (computing an arbitrary symmetric function) or THR-gate (computing an arbitrary linear threshold function) that is fed by S AND gates. Such circuits were considered in early influential work on unconditional derandomization of Luby, Veličković, and Wigderson [31], who gave the first non-trivial PRG with seed length $2^{O(\sqrt{\log(S/\varepsilon)})}$ that ε-fools these circuits.

In this work we obtain the first strict improvement of [31]'s seed length: we construct a PRG that ε-fools size-S $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AND}$ circuits over $\{0,1\}^n$ with seed length

$$2^{O(\sqrt{\log S})} + \mathrm{polylog}(1/\varepsilon),$$

an exponential (and near-optimal) improvement of the ε-dependence of [31]. The above PRG is actually a special case of a more general PRG which we establish for constant-depth circuits containing multiple SYM or THR gates, including as a special case $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0$ circuits. These more general results strengthen previous results of Viola [47] and essentially strengthen more recent results of Lovett and Srinivasan [30].

Our improved PRGs follow from improved correlation bounds, which are transformed into PRGs via the Nisan–Wigderson "hardness versus randomness" paradigm [37]. The key to our improved correlation bounds is the use of a recent powerful *multi-switching* lemma due to Håstad [21].

2012 ACM Subject Classification Theory of computation → Pseudorandomness and derandomization

Keywords and phrases Pseudorandom generators, correlation bounds, constant-depth circuits

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2018.56

1 Introduction

Depth-2 circuits which have a SYM or THR gate at the output and AND gates (of arbitrary fan-in) adjacent to the input variables are central objects of interest in concrete complexity, lying at the boundary of our understanding for many benchmark problems such as lower bounds, learning, and pseudorandomness. The class of SYM ∘ AND circuits has received

[1] Supported by NSF grants CCF-1420349 and CCF-1563155.
[2] Supported by NSF grant CCF-1563122. Part of this research was done during a visit to Columbia University.

© Rocco A. Servedio and Li-Yang Tan;
licensed under Creative Commons License CC-BY
Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2018).
Editors: Eric Blais, Klaus Jansen, José D. P. Rolim, and David Steurer; Article No. 56; pp. 56:1–56:20
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

much attention even in the restricted case of polylog(n) bottom fan-in (such circuits are also known as SYM^+ circuits) because of the well-known connection with the complexity class ACC^0 [52, 4, 10], a connection that is at the heart of recent breakthrough circuit lower bounds against ACC^0 [50, 34]. Another well-studied subclass, corresponding to the special case where the SYM gate computes the parity of its inputs, is the class of S-sparse polynomials over $\mathbb{F}_2$, which have been intensively studied in a wide range of contexts such as learning [42, 8, 9], approximation and interpolation [26, 18, 40], deterministic approximate counting [16, 27, 31], and property testing [13, 14]. Turning to THR gates (which compute an arbitrary linear threshold function of their inputs) as the top gate, the class of $\mathsf{THR} \circ \mathsf{AND}$ circuits of size-S is easily seen to contain the class of S-sparse polynomial threshold functions over $\{0,1\}^n$. This class, and special cases of it such as low-degree polynomial threshold functions, has also been intensively studied in complexity theory, learning theory, and derandomization, see e.g. [33, 17, 28, 24, 38, 32, 12, 15, 25, 11] and many other works. In this work we focus on *pseudorandom generators* for these $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AND}$ circuits.

In 1993 Luby, Veličković, and Wigderson [31] gave the first pseudorandom generators for these depth-2 circuits. As we shall discuss in detail below, this result was subsequently extended in various ways by different authors, but prior to the present work no strict improvement of Theorem 1 was known for the class of circuits that it addresses.

▶ **Theorem 1** (Luby–Veličković–Wigderson 1993). *There is a PRG with seed length $2^{O(\sqrt{\log(S/\varepsilon)})}$ that ε-fools the class of size-S $\mathsf{SYM} \circ \mathsf{AND}$ circuits over $\{0,1\}^n$. The same is true for the class of size-S $\mathsf{THR} \circ \mathsf{AND}$ circuits.*[3]

The main contribution of the present work is an exponential improvment of Theorem 1's dependence on ε, giving the first strict improvement of the [31] seed length:

▶ **Theorem 2** (Our main result). *There is a PRG with seed length $2^{O(\sqrt{\log S})} + \mathrm{polylog}(1/\varepsilon)$ that ε-fools the class of size-S $\mathsf{SYM} \circ \mathsf{AC}^0$ circuits. The same is true for $\mathsf{THR} \circ \mathsf{AC}^0$ circuits.*

Theorem 2 improves on a result of Viola [47] which, improving on [31], gave a $2^{O(\sqrt{\log(S/\varepsilon)})}$-seed-length PRG for size-S $\mathsf{SYM} \circ \mathsf{AC}^0$ circuits. The [47] PRG combines correlation bounds against $\mathsf{SYM} \circ \mathsf{AC}^0$ circuits with the Nisan–Wigderson "hardness versus randomness" paradigm, which yields pseudorandom generators from correlation bounds; we similarly prove Theorem 2 by establishing improved correlation bounds and using the Nisan–Wigderson paradigm.

Near-optimal hardness-to-randomness conversion. A major theme in computational complexity over the the last several decades, dating back to the seminal works of [44, 51, 5, 35, 37], has been that *computational hardness* can be converted into *pseudorandomness*. This insight is at the heart of essentially all unconditional pseudorandom generators, and motivates the goal of understanding when and how this conversion can be carried out in a quantitatively optimal manner. With this perspective in mind, we observe that the dependence on ε in Theorem 2 is optimal up to polynomial factors, and as we discuss in Section 1.3, achieving better dependence on S even for the special case of $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AND}$ circuits would require groundbreaking new lower bounds against low-degree $\mathbb{F}_2$ polynomials and ACC^0 circuits. Hence Theorem 2 achieves a near-optimal hardness-to-randomness conversion for $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0$ circuits; the seed length of our PRG is essentially the best possible given current state-of-the-art correlation bounds and circuit lower bounds.

[3] [31] does not actually consider $\mathsf{THR} \circ \mathsf{AND}$ circuits, but as we discuss later their arguments also apply to this class.

The exponential improvement in $1/\varepsilon$ over [31]'s seed length translates immediately into significantly improved deterministic approximate counting and deterministic search algorithms for $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0$ circuits, two basic algorithmic tasks in unconditional derandomization[4] (see e.g. [2] for formal definitions of these tasks and a discussion of how PRGs yield deterministic algorithms for them).

In the rest of this introduction we provide background and context for our results and explain the main ingredients that underlie them.

1.1 Prior PRGs and correlation bounds for $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0$

As mentioned above, the first results on PRGs for $\mathsf{SYM} \circ \mathsf{AND}$ circuits were given in early influential work of Luby, Veličković, and Wigderson [31], who constructed a PRG that ε-fools size-S $\mathsf{SYM} \circ \mathsf{AND}$ circuits over n variables with seed length $2^{O(\sqrt{\log(S/\varepsilon)})}$. The work of [31] employed ideas similar to those in the "hardness versus randomness" paradigm of [37], which subsequently came to be well understood as a versatile technique for constructing pseudorandom generators from correlation bounds.

A number of years later, with the [37] framework in hand, Viola [47] made the useful observation that correlation bounds against the larger class of $\mathsf{SYM} \circ \mathsf{AND} \circ \mathsf{OR}$ circuits translate to PRGs for $\mathsf{SYM} \circ \mathsf{AND}$ circuits in a "black-box" manner via [37], and the same is true when the top gate is THR instead of SYM. (Informally, the [37] translation "costs" two layers of depth: with typical parameter settings, it yields PRGs for a class $\mathcal{C}$ from correlation bounds against $\mathcal{C} \circ \mathsf{ANY}_{\log n}$ circuits, where an ANY_t gate computes an arbitrary t-variable Boolean function. By rewriting the $\mathsf{ANY}_{\log n}$ gate as a CNF, it is possible to collapse the two adjacent layers of AND gates, yielding Viola's observation.) Roughly speaking, in this translation from correlation bounds against $\mathcal{C} \circ \mathsf{ANY}_{\log n}$ to PRGs that ε-fool $\mathcal{C}$,

- the larger the $\mathcal{C} \circ \mathsf{ANY}_{\log n}$ circuits for which the correlation bound holds, the better (smaller) is the PRG's seed length for fooling size-S functions in $\mathcal{C}$; and
- the smaller the advantage over random guessing that the correlation bound establishes, the better (smaller) is the PRG's seed length's dependence on the fooling parameter ε.

Motivated by this template, [47] established $n^{-\Omega(\log n)}$ correlation bounds against $\mathsf{SYM} \circ \mathsf{AC}^0$ circuits of size $n^{\Omega(\log n)}$. This translates (see Appendix B) into a PRG with seed length $2^{O(\sqrt{\log(S/\varepsilon)})}$ for size-S $\mathsf{SYM} \circ \mathsf{AC}^0$ circuits over $\{0,1\}^n$, matching the seed length achieved by [31] but for a larger class of circuits (and also with a simpler and more modular proof). While [47] does not explicitly discuss THR gates, its proof like that of [31] also goes through for $\mathsf{THR} \circ \mathsf{AC}^0$ as remarked in the earlier footnote.

Subsequent work of [30] established a strong correlation bound of $\exp(-\Omega(n^{1-o(1)}))$ against $\mathsf{SYM} \circ \mathsf{AC}^0$ and a correlation bound of $\exp(-\Omega(n^{1/2-o(1)}))$ against $\mathsf{THR} \circ \mathsf{AC}^0$, but in both cases only for such circuits of size $n^{O(\log\log n)}$. Via the Nisan–Wigderson framework [37] this translates into a PRG with seed length $2^{O(\log S/\log\log S)} + \mathrm{polylog}(1/\varepsilon)$ for size-S $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0$ circuits over $\{0,1\}^n$; while this is a very good dependence on ε, it comes at the cost of a significantly worse dependence on the circuit size S. Thus both the seed length and correlation bounds of [30] are incomparable to those of [31, 47]; see Table 1. We also note that the recent work of Sakai et al. [41] established a range of tradeoffs between circuit size and correlation bounds for (a generalization of) $\mathsf{SYM} \circ \mathsf{AC}^0$ circuits, though for the circuit sizes presented in Table 1 their results are not as strong as the results of [47, 30] listed there.

[4] Indeed, the work of [31] was explicitly motivated by deterministic approximate counting of S-sparse $\mathbb{F}_2$ polynomials; see the abstract of [31].

Table 1 Correlation bounds against $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0{}_d$ circuits and the PRGs that follow via the [37] paradigm. In all cases the "hard function" is the RW function that is defined in (3) and was first considered by Razborov and Wigderson [39]. For a given row, a circuit size of s and a correlation bound of α means that every size-s circuit of the stated type agrees with the n-variable RW function on at most $\frac{1}{2} + \alpha$ fraction of inputs. For each row, see Appendix B for a derivation of how the final column (seed length for a Nisan–Wigderson based PRG) follows from the earlier columns via [37].

	Circuit type	Circuit size S	Correlation bound	PRG seed length
[47]	$\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0{}_d$	$n^{c_d \log n}$	$n^{-c_d \log n}$	$2^{O(\sqrt{\log(S/\varepsilon)})}$
[30]	$\mathsf{SYM} \circ \mathsf{AC}^0{}_d$	$n^{c_d \log\log n}$	$\exp(-n^{1-o(1)})$	$2^{O\left(\frac{\log S}{\log\log S}\right)} + (\log(1/\varepsilon))^{2+o(1)}$
[30]	$\mathsf{THR} \circ \mathsf{AC}^0{}_d$	$n^{c_d \log\log n}$	$\exp(-n^{1/2-o(1)})$	$2^{O\left(\frac{\log S}{\log\log S}\right)} + (\log(1/\varepsilon))^{4+o(1)}$
This work	$\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0{}_d$	$n^{c \log n}$	$\exp(-\Omega(n^{0.499}))$	$2^{O(\sqrt{\log S})} + (\log(1/\varepsilon))^{4.01}$

(We further note that other incomparable results have been achieved in separate lines of work on pseudorandom generators for degree-d polynomial threshold functions [12, 32, 25] and degree-d $\mathbb{F}_2$ polynomials [6, 7, 29, 49], which correspond to $\mathsf{THR} \circ \mathsf{AND}_d$ and $\mathsf{PAR} \circ \mathsf{AND}_d$ circuits respectively. The seed lengths of these PRGs all have an exponential dependence on d, and thus do not yield non-trivial results for general poly(n)-size $\mathsf{THR} \circ \mathsf{AND}$ or $\mathsf{PAR} \circ \mathsf{AND}$ circuits. For constant d, the [29, 49] PRGs for $\mathsf{PAR} \circ \mathsf{AND}_d$ circuits achieve optimal seed length, while the [32, 25] PRGs for $\mathsf{THR} \circ \mathsf{AND}_d$ have seed length poly$(1/\varepsilon) \cdot \log n$.)

1.2 Our main technical contribution: New correlation bounds against $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0$ circuits

The technical heart of our main result is a new exponential correlation bound against $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0$ circuits of size $n^{\Omega(\log n)}$:

▶ **Theorem 3.** *There is an absolute constant $\tau > 0$ and an explicit* poly(n)*-time computable function $H : \{0,1\}^n \to \{0,1\}$ with the following property: for any constant d, for n sufficiently large it is the case that for any n-variable circuit C of size $n^{\tau \log n}$ and depth d with a* SYM *or* THR *gate at the top, we have*

$$\Pr_{\boldsymbol{x} \leftarrow \{0,1\}^n}[H(\boldsymbol{x}) = C(\boldsymbol{x})] \leq \frac{1}{2} + \exp(-\Omega(n^{0.499})).$$

Theorem 3 strictly improves on the correlation bound provided by Theorem 4 of [47], as it establishes correlation bounds for the same class of $n^{\Omega(\log n)}$-size circuits, but gives a much smaller $\exp(-\Omega(n^{0.499}))$ upper bound on the correlation rather than $n^{-\Omega(\log n)}$. As described in Appendix B, our PRG result for $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0$ (Theorem 2) follows directly from Theorem 3 via the Nisan–Wigderson framework. In Section 1.4 we give an overview of the ideas that underlie our new correlation bound.

Correlation bounds and PRGs for constant-depth circuits with multiple SYM or THR gates. The main correlation bound and PRG of [47] are actually for $n^{c_d \log n}$-size depth-d circuits with $c_d(\log n)^2$ many SYM gates, and similarly the main result of [30] is a correlation bound for constant-depth circuits with $n^{1-o(1)}$ many SYM gates or $n^{1/2-o(1)}$ many THR gates. Our results similarly extend to constant-depth circuits with multiple SYM or THR gates. Our most general correlation bound is the following:

▶ **Theorem 4.** *There is an absolute constant $\tau > 0$ and an explicit* $\mathrm{poly}(n)$*-time computable function $H : \{0,1\}^n \to \{0,1\}$ with the following property: for any constant d, for n sufficiently large, any n-variable circuit C of size $n^{\tau \log n}$ and depth d containing $n^{0.249}$ many* SYM *or* THR *gates (the circuit is allowed to contain both types of gates) satisfies*

$$\Pr_{\boldsymbol{x} \leftarrow \{0,1\}^n}[H(\boldsymbol{x}) = C(\boldsymbol{x})] \leq \frac{1}{2} + \exp(-\Omega(n^{0.249})).$$

Via the Nisan–Wigderson framework, Theorem 4 immediately yields the following, which is our most general PRG result:

▶ **Corollary 5.** *For some sufficiently small absolute constant $c > 0$, there is a PRG with seed length $2^{O(\sqrt{\log S})} + \mathrm{polylog}(1/\varepsilon)$ that ε-fools the class of size-S constant-depth circuits that contain $2^{\sqrt{c \log S}}$ many* SYM *or* THR *gates.*

This strictly improves the main [47] PRG (Theorem 1 of [47]), which achieves seed length $2^{O(\sqrt{\log(S/\varepsilon)})}$ for size-S constant-depth circuits that contain $O((\log S)^2)$ many SYM or THR gates. We prove Theorem 4 in Appendix C.

1.3 Barriers to further progress: correlation bounds for $\mathbb{F}_2$ polynomials and ACC^0 lower bounds

In this section we outline why achieving better dependence on S will require groundbreaking new correlation bounds or circuit lower bounds.

The seminal work of Babai, Nisan, and Szegedy [3] gave an explicit function and established that it has exponentially small correlation $\exp(-\Omega(n/4^d\, d))$ with any n-variable $\mathbb{F}_2$ polynomials of degree d (see Theorem 3 of [48]). This result (and the multiparty communication-based techniques underlying it) have had far-reaching consequences in complexity theory; 25 years later, improving on this correlation bound remains a prominent open problem. In particular, even achieving correlation bounds of the form $\frac{1}{2} + n^{-1}$ against polynomials of degree $\log n$ with respect to any explicit distribution $\mathcal{D}$ would constitute a significant breakthrough (see e.g. "Open Question 1" in Viola's excellent survey [48]). Since every degree-d polynomial is s-sparse for $s = \binom{n}{d}$, this is clearly a special case of obtaining $\frac{1}{2} + n^{-1}$ correlation bounds against polynomials of sparsity $s = \binom{n}{\log n}$. Via a standard connection between PRGs and correlation bounds (see e.g. Proposition 3.1 of [49]), an improvement in the dependence on S in Theorem 2 to $2^{o(\sqrt{\log S})} + \mathrm{polylog}(1/\varepsilon)$, even for the special case of S-sparse $\mathbb{F}_2$ polynomials, would immediately yield exponentially-small correlation bounds against s-sparse $\mathbb{F}_2$ polynomials for $s = n^{\omega(\log n)} \gg \binom{n}{\log n}$ with respect to an explicit distribution. (We remark that the same correlation bounds are also open for the class of degree $\log n$ polynomial threshold functions, and hence the same barrier applies to improving the S-dependence of PRGs for size-S $\mathsf{THR} \circ \mathsf{AND}$ circuits.)

Further improvements of Theorem 2 would have even more dramatic consequences. Classical "depth-compression" results of Yao [52] and Beigel and Tauri [4] (see also [10]) show that every size-s depth-d ACC^0 circuit can be computed by a size-S $\mathsf{SYM} \circ \mathsf{AND}$ circuit where $S = \exp((\log n)^{O_d(1)})$. Improving the seed length of Theorem 2 for the class of size-S $\mathsf{SYM} \circ \mathsf{AND}$ circuits to $2^{(\log S)^{o(1)}}$ (even for constant ε) would therefore separate NP from ACC^0, a significant strengthening of Williams's celebrated separation of NEXP from ACC^0 [50].

1.4 The high-level structure of our correlation bound argument

We recall the "bottom-up" approach to proving correlation bounds via the method of random restrictions. This approach dates back to the classic correlation bounds between Parity

and AC^0 of [1] and [20]; in particular, the relevant prior works of [47, 30] also operate within this framework.

Fix a hard function H, and let F be any function belonging to a given class $\mathcal{F}$ of Boolean functions (in our case $\mathcal{F}$ is the class of $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0$ circuits of size $n^{\Omega(\log n)}$). Our goal is to show that F has small *correlation* with H, i.e. that $\Pr_{\boldsymbol{x} \leftarrow \{0,1\}^n}[F(\boldsymbol{x}) = H(\boldsymbol{x})] \leq \frac{1}{2} + \alpha$ for some small α where $\boldsymbol{x}$ is uniform over $\{0,1\}^n$. This can be achieved by designing a *fair* distribution $\mathcal{R}$ over random restrictions that satisfies the following two competing requirements. (A distribution $\mathcal{R}$ over restrictions is said to be fair if first drawing a restriction $\boldsymbol{\rho} \leftarrow \mathcal{R}$ and then filling in all $*$'s to independent uniform values from $\{0,1\}$ results in a uniform random string from $\{0,1\}^n$.)

(1) Approximator (F) simplifies: With high probability $1 - \gamma_{\mathrm{SL}}$ over $\boldsymbol{\rho} \leftarrow \mathcal{R}$, F "collapses" when it is hit by $\boldsymbol{\rho}$, meaning that $F \restriction \boldsymbol{\rho} \in \mathcal{F}_{\mathrm{simple}}$ for some class $\mathcal{F}_{\mathrm{simple}} \subseteq \mathcal{F}$. Looking ahead, in our case

$$\mathcal{F}_{\mathrm{simple}} = \Big\{ \{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AND}_{k:=0.0005 \log m} \text{ circuits} \Big\}$$

where $m \approx \sqrt{n}$ and AND_k denotes the class of fan-in k AND gates. A collapse to this $\mathcal{F}_{\mathrm{simple}}$ is useful for us because there are efficient multiparty communication protocols for functions computable by $\mathcal{F}_{\mathrm{simple}}$ (due to [22] when the top gate is SYM and to [36] when it is THR).

(2) Target (H) retains structure: With high probability $1 - \gamma_{\mathrm{target}}$ over $\boldsymbol{\rho} \leftarrow \mathcal{R}$, the restricted hard function $H \restriction \boldsymbol{\rho}$ "retains structure", in the sense that it has small correlation with every function in $\mathcal{F}_{\mathrm{simple}}$. In our case our notion of structure will be that $H \restriction \boldsymbol{\rho}$ "contains a perfect copy of" the generalized inner product function:

$$\mathrm{GIP}_{m/2,k+1}(x) := \bigoplus_{i=1}^{m/2} \bigwedge_{j=1}^{k+1} x_{i,j},$$

where m and k are the same m and k as above.

Suppose we have such a fair distribution $\mathcal{R}$ over random restrictions satisfying (1) and (2) above. The remaining step in the argument is to show the following: (3) for any ρ such that both of the above happen (approximator simplifies and the hard function retains structure), $F \restriction \rho$ and $H \restriction \rho$ have small correlation, i.e. they agree on at most $\frac{1}{2} + \gamma_{\mathrm{corr}}$ fraction of all inputs. As in the previous works of [47, 30], the fact that $\mathcal{F}_{\mathrm{simple}}$ and $\mathrm{GIP}_{m/2,k+1}$ have small correlation follows from a celebrated theorem of Babai, Nisan, and Szegedy [3] lower bounding the multiparty communication complexity of $\mathrm{GIP}_{m,k+1}$.

It is straightforward to see that items (1)–(3) above establish a correlation bound of

$$\Pr_{\boldsymbol{x} \leftarrow \{0,1\}^n}[\, F(\boldsymbol{x}) = H(\boldsymbol{x}) \,] \leq \frac{1}{2} + \gamma_{\mathrm{SL}} + \gamma_{\mathrm{target}} + \gamma_{\mathrm{corr}}.$$

The goal is therefore to carry out the above with $\max\{\gamma_{\mathrm{SL}}, \gamma_{\mathrm{target}}, \gamma_{\mathrm{corr}}\}$ as small as possible for the class $\mathcal{F}$ of $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0{}_d$ circuits of as large a size as possible. As indicated earlier, for any constant d and circuits of size up to $s = n^{\tau \log n}$ we achieve $\max\{\gamma_{\mathrm{SL}}, \gamma_{\mathrm{target}}, \gamma_{\mathrm{corr}}\} = \exp(-\Omega(n^{0.499}))$.

We upper bound γ_{SL}, γ_{target}, and γ_{corr} in Sections 2, 3, and 4 respectively. Some useful preliminaries are given in Appendix A.

1.5 How this work differs from [47, 30]: improved depth reduction

A simple observation (due to [19]) that is used in both [47, 30] and in our work as well is the fact that a symmetric function of depth-k decision trees can be simulated by a (different) symmetric function of width-k AND's, and likewise for a threshold function of depth-k decision trees. (See Fact 1 for a precise statement.) Consequently we can think of $\mathcal{F}_{\text{simple}}$ as $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{DT}_k$ rather than $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AND}_k$ (where DT_k denotes the class of decision trees of depth k), and for depth reduction it suffices to prove that a family of s many AC^0 circuits collapses to a family of small-depth decision trees with high probability under a random restriction. This is exactly what is shown by *switching lemmas*.

The loss in the previous works of [47, 30] is due to the switching lemmas they use and the limitations of these switching lemmas. [47] uses the standard [20] switching lemma:

▶ **Theorem 6** (Håstad's switching lemma). *Let F be computed by a depth-2 circuit with bottom fan-in w. Then*

$$\Pr_{\boldsymbol{\rho} \leftarrow \mathcal{R}_p} \left[F \restriction \boldsymbol{\rho} \text{ is not a depth-}t \text{ decision tree} \right] \leq (5pw)^t.$$

This failure probability of $(5pw)^t$ cannot be made exponentially small in our setting: since correlation bounds strong enough to be useful for the [37] framework are not known for $\mathsf{SYM} \circ \mathsf{AND}_{\omega(\log n)}$ (see "Open Question 1" of [48]) the value of t has to has to be taken to be at most $k = O(\log n)$, and moreover p certainly has to be $\gg 1/n$ (since taking $p = 1/n$ would leave only a constant number of coordinates alive, and $H \restriction \boldsymbol{\rho}$ would not "retain structure" in the sense of containing a copy of $\mathrm{GIP}_{m/2,k+1}$). Indeed, [47] applies Theorem 6 with $p = n^{-\Theta(1)}$ in order to make the failure probability as small as $n^{-\Omega(\log n)}$, and this is why [47] only achieves quasi-polynomial correlation bounds $n^{-\Omega(\log n)}$.

Faced with this obstacle, instead of using the standard [20] switching lemma, [30] reverts to the earlier "multi-switching lemma" of [1] which applies to a collection of depth-2 circuits rather than a single such circuit. The [1] multi-switching lemma, stated below, *does* achieve exponentially small failure probability, but is only able to handle collections of $n^{O(\log \log n)}$ many k-DNFs, for $k = O(\log \log n)$. Recall that a restriction tree T is like a decision tree except that leaves do not have labels associated with them (so each root-to-leaf path is a restriction). The distribution μ_T corresponds to the distribution over restrictions obtained by making a random walk from the root of T.

▶ **Theorem 7** (Ajtai's switching lemma [1]). *Let $\mathcal{F} = \{F_1, \ldots, F_s\}$ be a family of s many DNFs over $x_1, \ldots, x_n$, each of width k. For any $t \geq 1$, there is a restriction tree T of height at most $nk(\log s)/(\log n)^t$ such that*

$$\Pr_{\boldsymbol{\rho} \leftarrow \mu_T} \left[F_i \restriction \boldsymbol{\rho} \text{ is not a } (\log n)^{10kt2^k}\text{-junta} \right] \leq 2^{-n/(2^{10k}(\log n)^t)}.$$

Hence [30] achieves exponentially small correlation bounds (the main point of their paper), but only against circuits of size $n^{O(\log \log n)}$.

The key new ingredient that we employ in this work is a recent powerful multi-switching lemma from [21]. (We note that [23] gives an essentially equivalent multi-switching lemma which we could also use.) Roughly speaking the [21] multi-switching lemma, whose precise statement we defer to Section 2 as it is somewhat involved, lets us achieve an exponentially small failure probability (like Ajtai's multi-switching lemma) of achieving a significantly more drastic simplification than Ajtai's multi-switching lemma (recall the doubly-exponential-in-k dependence on the junta size in Theorem 7). This quantitative improvement in depth reduction translates into our stronger correlation bounds.

1.6 Relation to [43]

We close this introduction by discussing the connection between this paper and recent concurrent work of the authors [43]. The high-level approaches of the two papers are fairly different: unlike the current paper, [43] does not use the Nisan–Wigderson hardness-versus-randomness paradigm (and does not establish any new correlation bounds); instead it establishes a derandomized version of the [21] multi-switching lemma and combines this with other ingredients to obtain its final PRG in a manner reminiscent of [2, 45].

The results of the two papers are also incomparable (briefly, [43] obtains significantly shorter seed length for significantly more restricted classes of functions). The first main result of [43] is an ε-PRG for the class of size-S depth-d AC^0 circuits with seed length $\log(S)^{d+O(1)} \cdot \log(1/\varepsilon)$. This is incomparable to the most closely related result of the present paper (Corollary 5, which gives a $2^{O(\sqrt{\log S})} + \mathrm{polylog}(1/\varepsilon)$ seed length PRG for AC^0 circuits augmented with polynomially many SYM or THR gates), since the [43] result gives a significantly better seed length but for the significantly more limited class of "un-augmented" constant-depth circuits (indeed, the [43] result does not apply to AC^0 circuits augmented even with a single SYM or THR gate). The second main result of [43] is an ε-PRG for the class of S-sparse $\mathbb{F}_2$ polynomials with seed length $2^{O(\sqrt{\log S})} \cdot \log(1/\varepsilon)$. Here too the seed length of [43] is shorter than that of the current paper (giving the optimal $\log(1/\varepsilon)$ dependence on ε as opposed to the $(\log(1/\varepsilon))^{4.01}$ of the current paper), but the result of [43] only holds for S-sparse $\mathbb{F}_2$ polynomials, which are a very restricted case of the $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0{}_d$ circuits which are handled in the current paper.

2 Ingredient (1): Simplifying the approximator

The main result of this section is the following:

▶ **Lemma 8.** *Let F be any $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0{}_d$ circuit of size $s = n^{\tau \log n}$. There is a fair distribution $\mathcal{R}$ over restrictions $\boldsymbol{\rho} \in \{0, 1, *\}^n$ such that the following holds: With probability $1 - \gamma_{\mathrm{SL}} = 1 - \exp(-\Omega_d(\sqrt{n/\log n}))$ over the draw of $\boldsymbol{\rho} \leftarrow \mathcal{R}$, it is the case that $F \restriction \boldsymbol{\rho}$ belongs to the class $\mathcal{F}_{simple} = \{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AND}_{k:=0.0005 \log m}$, where $m = \Theta(\sqrt{n/\log n})$.*

The recent "multi-switching lemma" of [21] is the main technical tool we use to establish Lemma 8. To state the [21] lemma we need some terminology. Let $\mathcal{G}$ be a family of Boolean functions. A restriction tree T is said to be a *common ℓ-partial restriction tree (RT) for* $\mathcal{G}$ if every $g \in \mathcal{G}$ can be expressed as T with depth-ℓ decision trees hanging off its leaves. (Equivalently, for every $g \in \mathcal{G}$ and root-to-leaf path π in T, we have that $g \restriction \pi$ is computed by a depth-ℓ decision tree.)

▶ **Theorem 9** ([21] multi-switching lemma)**.** *Let $\mathcal{F} = \{F_1, \ldots, F_s\}$ be a collection of depth-2 circuits with bottom fan-in w. Then for any $t \geq 1$,*

$$\Pr_{\boldsymbol{\rho}' \leftarrow \mathcal{R}_p} \left[\mathcal{F} \restriction \boldsymbol{\rho}' \text{ does not have a common } (\log s)\text{-partial RT of depth} \leq t \right] \leq s(24pw)^t.$$

Theorem 9 is the main tool we use to simplify any $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0$ circuit down to an $\mathcal{F}_{\mathrm{simple}}$-circuit. Conceptually, we think of this transformation as being done in three steps:

1. (Main step) Apply a random restriction $\boldsymbol{\rho}' \leftarrow \mathcal{R}_p$ to convert a $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0$ circuit into a decision tree with a $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{DT}$ circuit at each leaf.
2. Observing that $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{DT} \equiv \{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AND}$, this is equivalent to a decision tree with a $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AND}$ circuit at each leaf.

3. Trim the fan-in of the AND gates in each $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AND}$ circuit (by increasing the depth of the decision tree). The last step in the draw of a random restriction $\boldsymbol{\rho}$ from the overall fair distribution $\mathcal{R}$ corresponds to a random walk down this final decision tree.

In the rest of this section we describe each of these steps in detail and thereby prove Lemma 8.

First (main) step. If g is a Boolean function and $\mathcal{C}$ is a class of circuits, we say that g is *computed by a* $(d, \mathcal{C})$*-decision tree* if g is computed by a decision tree of depth d (with a single Boolean variable at each internal node as usual) in which each leaf is labeled by a function from $\mathcal{C}$. We require the following corollary of Theorem 9:

▶ **Corollary 10.** *Let G be any Boolean function and G be a gate computing G, and let F be a $\mathsf{G} \circ \mathsf{AC}^0{}_d$ circuit of size s . Then for $p = \frac{1}{48}(48 \log s)^{-(d-1)}$ and any $t \geq 1$,*

$$\Pr_{\boldsymbol{\rho}' \leftarrow \mathcal{R}_p} \left[F \restriction \boldsymbol{\rho}' \text{ is not computed by a } (2^d t, \mathsf{G} \circ \mathsf{DT}_{\log s})\text{-decision tree} \right] \leq s \cdot 2^{-t}.$$

Proof. We may assume without loss of generality that the depth-$(d+1)$ circuit F is *layered*, meaning that for any gate g it contains, every directed path from an input variable to g has the same length (converting an unlayered circuit to a layered one increases its size only by a factor of d, which is negligible for our purposes). Let s_i denote the number of gates in layer i (at distance i from the inputs), so $s = s_1 + \cdots + s_d$.

We begin by trimming the bottom fan-in of F: applying Theorem 9 with $\mathcal{F}$ being the s_1 many bottom layer gates of F (viewed as depth-2 circuits of bottom fan-in $w = 1$) and $p_0 := 1/48$, we get that

$$\Pr_{\boldsymbol{\rho}_0 \leftarrow \mathcal{R}_{p_0}} \left[F \restriction \boldsymbol{\rho}_0 \text{ is not computed by a } (t, G \circ \mathsf{AC}^0(\text{depth } d, \text{ bottom fan-in } \log s))\text{-decision tree} \right] \leq s_1 \cdot 2^{-t}.$$

Let $F^{(0)}$ be any good outcome of the above, a $(t, G \circ \mathsf{AC}^0(\text{depth } d, \text{ bottom fan-in } \log s))$-decision tree. Note that there are at most 2^t many $\mathsf{AC}^0(\text{depth } d, \text{ fan-in } \log s)$ circuits at the leaves of the depth-t decision tree. Applying Theorem 9 to each of them with $p_1 := 1/(48 \log s)$ (and the 't' of Theorem 9 being $2t$) and taking a union bound over all 2^t many of them, we get that

$$\Pr_{\boldsymbol{\rho}_1 \leftarrow \mathcal{R}_{p_1}} \left[F^{(0)} \restriction \boldsymbol{\rho}_1 \text{ is not a } (t + 2t, G \circ \mathsf{AC}^0(\text{depth } d-1, \text{ fan-in } \log s))\text{-decision tree} \right]$$
$$\leq s_2 \cdot 2^{-2t} \cdot 2^t = s_2 \cdot 2^{-t}.$$

Repeat with $p_2 = \ldots = p_{d-1} := 1/(48 \log s)$, each time invoking Theorem 9 with its 't' being the one more than the current depth of the decision tree . The claim then follows by summing the $s_1 2^{-t}$, $s_2 2^{-t}, \ldots, s_d 2^{-t}$ failure probabilities over all d stages and the fact that

$$\prod_{j=0}^{d-1} p_i = \frac{1}{48} \cdot \frac{1}{(48 \log s)^{d-1}} = p.$$ ◀

Second step: From $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{DT}$ to $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AND}$. We recall the following fact from [19]:

▶ **Fact 1.** *Every $\mathsf{SYM}_s \circ \mathsf{DT}_{\log s}$ function (resp. $\mathsf{THR}_s \circ \mathsf{DT}_{\log s}$) can be computed by a $\mathsf{SYM}_{s^2} \circ \mathsf{AND}_{\log s}$ (resp. $\mathsf{THR}_{s^2} \circ \mathsf{AND}_{\log s}$) circuit.*

(This is an easy consequence of the fact that any decision tree may be viewed as a DNF whose terms corresponds to the paths to 1-leaves, and that this DNF has the property that any input assignment makes at most one term true.) Applying Fact 1 and choosing $t = m/2^{d+1}$ in Corollary 10, (where $m = \Theta(\sqrt{n/\log n})$ will be defined precisely in the next section), we get the following special case of Corollary 10:

▶ **Corollary 11.** *Let F be a $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0{}_d$ circuit of size $s = n^{\tau \log n}$. Then for $p = \frac{1}{48}(48 \log s)^{-(d-1)}$,*

$$\begin{aligned}
&\Pr_{\boldsymbol{\rho}' \leftarrow \mathcal{R}_p}\left[\, F \restriction \boldsymbol{\rho}' \text{ is not computed by a } (m/2, \{\mathsf{SYM}_{s^2}, \mathsf{THR}_{s^2}\} \circ \mathsf{AND}_{\log s})\text{-decision tree} \,\right] \\
&\le s \cdot 2^{-t} = s \cdot 2^{-m/2^{d+2}} \\
&= \exp(-\Omega_d(\sqrt{n/\log n})) \;:=\; \gamma_{\mathrm{SL}}. \qquad (1)
\end{aligned}$$

Third step: Trimming to reduce bottom fan-in. The $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AND}$ circuits hanging off the leaves of our decision tree have bottom fan-in at most $\log s$, but we will need them to have fan-in at most k in order to invoke the [3] lower bound later. At each leaf ℓ we achieve this smaller fan-in by identifying a set (call it S_ℓ) of additional variables and restricting them in all possible ways; we argue that every fixing of the variables in S_ℓ gives the desired upper bound of k on the bottom-AND fan-in. We use a probabilistic argument to establish the existence of the desired set S_ℓ (this is important because in the next section we will need each S_ℓ to satisfy an additional property, and the probabilistic argument makes it easy to achieve this).

Let us write "$\boldsymbol{L} \subseteq_q X$" to indicate that $\boldsymbol{L}$ is a subset of X that is randomly chosen by independently including each element of X with probability q. We will use the following easy result:

▶ **Fact 2.** *Let $\{C_1, \ldots, C_{s^2}\}$ be a collection of subsets of $[n]$ where each $|C_i| \le w$. Then for $\boldsymbol{L} \subseteq_q [n]$ and $k \le w$, we have*

$$\Pr_{\boldsymbol{L} \subseteq_q [n]}\left[\, \exists\, i \in [s^2] \text{ such that } |C_i \cap \boldsymbol{L}| > k \,\right] \le s^2 \binom{w}{k} q^k.$$

Recall that $s = n^{\tau \log n}$ where $\tau > 0$ is a small absolute constant to be specified later and that $k = 0.0005 \log n$. We set

$$q := \frac{k}{e \log s} \cdot 2^{-(3 \log s)/k} = \frac{1}{\Theta(\log n)} \cdot n^{-\Theta(1)\cdot\tau} < n^{-0.01},$$

where the last inequality holds for a suitably small choice of the constant τ. Observe that q is chosen so as to ensure

$$s^2 \binom{\log s}{k} q^k \le 2^{2 \log s} \left(\frac{e \log s}{k} \cdot q\right)^k = \frac{1}{s} \ll 1. \qquad (2)$$

Fix T to be an $(m/2, \{\mathsf{SYM}_{s^2}, \mathsf{THR}_{s^2}\} \circ \mathsf{AND}_{\log s})$-decision tree as given by Corollary 11. At each leaf ℓ of T, draw a set $\boldsymbol{L}(\ell) \subseteq_q [n]$ and let $\mathbf{S}_\ell$ be $([n] \setminus \mathrm{fixed}(\ell)) \setminus \boldsymbol{L}(\ell)$, where $\mathrm{fixed}(\ell) \subseteq [n]$ is the subset of variables that are fixed on the root-to-ℓ path in T. By Fact 2 and (2), at each leaf ℓ it is the case that with probability at least $1 - 1/s$ over the random draw of $\boldsymbol{L}(\ell)$, every extension of the root-to-ℓ path in T that additionally fixes all the variables in $\mathbf{S}_\ell$ collapses the $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AND}_{\log s}$ circuit that was at ℓ in T down to a

$\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AND}_k$ circuit. We say that such an outcome of $\boldsymbol{L}(\ell)$ is a *good* outcome (we will refer back to this notion in the next section).

In summary, the above discussion establishes Lemma 8, where the fair distribution $\mathcal{R}$ corresponds to

(a) first drawing $\boldsymbol{\rho}' \leftarrow \mathcal{R}_p$,

(b) then walking down a random root-to-leaf path π in the resulting depth-$(m/2)$ decision tree given by Corollary 11,

(c) and then finally, at the resulting leaf ℓ, choosing a random assignment to the variables in the set S_ℓ that corresponds to $L(\ell)$, where $L(\ell)$ is a good outcome of the random variable $\boldsymbol{L}(\ell) \subseteq_q [n]$. (Note that the randomness over $\boldsymbol{L}(\ell)$ is not part of the random draw of $\boldsymbol{\rho} \leftarrow \mathcal{R}$; all we require is the existence of a good $L(\ell)$.)

Based on our discussion thus far each $L(\ell)$ may be fixed to be any good outcome of $\boldsymbol{L}(\ell)$; we will give an additional stipulation on $L(\ell)$ in Remark 3.

3 Ingredient (2) (target retains structure): GIP ∘ PAR under random restrictions

Like [47, 30], our hard function will be the generalized inner product function composed with parity:

$$\mathrm{RW}_{m,k,r}(x) = \bigoplus_{i=1}^{m} \bigwedge_{j=1}^{k+1} \bigoplus_{\ell=1}^{r} x_{i,j,\ell}. \tag{3}$$

This function was introduced by Razborov and Wigderson [39] to show $n^{\Omega(\log n)}$ lower bounds against depth-3 threshold circuits with AND gates at the bottom layer. We will set

$$m = r = \sqrt{n/(k+1)} \quad (\text{recall that } k = 0.0005 \log m).$$

Note that $m = r = \Theta(\sqrt{n/\log n})$ and $k = \Theta(\log n)$. Given parameters m', k', r', we say that a function $g : \{0,1\}^n \to \{0,1\}$ *contains a perfect copy of* $\mathrm{RW}_{m',k',r'}$ if there is a restriction κ such that $(g \restriction \kappa)(x) = b \oplus \bigoplus_{i=1}^{m'} \bigwedge_{j=1}^{k+1} \left(b_{i,j} \bigoplus_{\ell=1}^{r'} x_{i,j,k}\right)$ for some bits $b, b_{i,j}$.

Roughly speaking, the motivation behind augmenting GIP with a layer of parities is to ensure that RW is resilient to random restrictions (i.e. that $\mathrm{RW} \restriction \boldsymbol{\rho}$ "remains complex", containing a copy of GIP with high probability after a suitable random restriction). In our setting we need that RW is resilient to a random restriction $\boldsymbol{\rho} \leftarrow \mathcal{R}$ for the fair distribution $\mathcal{R}$ from Lemma 8; we establish this in the rest of this section.

▶ **Proposition 1.** *Consider the space of formal variables of* $\mathrm{RW}_{m,k,r} : \{0,1\}^n \to \{0,1\}$*:*

$$X = \left\{x_{i,j,t} \colon (i,j,t) \in [m] \times [k+1] \times [r]\right\}, \quad |X| = m(k+1)r := n.$$

Then for $p = \frac{1}{48}(48 \log s)^{-(d-1)}$ *(as in Corollary 11),*

$$\underset{\boldsymbol{L} \subseteq_p X}{\boldsymbol{Pr}} \left[\exists\, (i,j) \colon \left|\{t \in [r] \colon x_{i,j,t} \in \boldsymbol{L}\}\right| < \frac{pr}{2}\right] \le m(k+1) \cdot \exp\left(-\Omega(pr)\right).$$

Proof. This follows directly from a standard multiplicative Chernoff bound and a union bound over all $(i,j) \in [m] \times [k+1]$. ◀

Recall that a random restriction $\boldsymbol{\rho}' \leftarrow \mathcal{R}_p$ can be thought of as being sampled by first drawing $\mathbf{K} \subseteq_p X$ and setting $\boldsymbol{\rho}_i$ to $*$ for each $i \in \mathbf{K}$, and then setting the coordinates of $\boldsymbol{\rho}'$ in $X \setminus \mathbf{K}$ according to a uniform random draw from $\{0,1\}^{X \setminus \mathbf{K}}$. Proposition 1 and the definition of RW thus yield the following:

▶ **Corollary 12.** *For $\boldsymbol{\rho}' \leftarrow \mathcal{R}_p$, for $p = \frac{1}{48}(48 \log s)^{-(d-1)}$, $\mathrm{RW}_{m,k,r}(x) \restriction \boldsymbol{\rho}'$ contains a perfect copy of*

$$\mathrm{RW}_{m,k,r'}(x) = \bigoplus_{i=1}^{m} \bigwedge_{j=1}^{k+1} \bigoplus_{t=1}^{r'} x_{i,j,t}, \qquad \textit{where} \quad r' = \frac{pr}{2}$$

with failure probability at most

$$\exp(-\Omega(pr)) = \exp\left(-\frac{\sqrt{n/\log n}}{(\Theta(\log s))^{d-1}}\right) := \gamma_{\mathrm{target}}. \tag{4}$$

Note that

$$r' = \frac{pr}{2} = \frac{\sqrt{n/\log n}}{(\Theta(\log s))^{d-1}} > n^{0.49},$$

where the inequality uses the fact that d is a constant and the fact that $s = n^{O(\log n)}$; we will use this later.

Corollary 12 states that with very high probability over $\boldsymbol{\rho}' \leftarrow \mathcal{R}_p$, the function $\mathrm{RW}_{m,k,r} \restriction \boldsymbol{\rho}'$ "does not simplify too much"; however we need $\mathrm{RW}_{m,k,r}$ to "not simplify too much" under a full random restriction drawn from $\mathcal{R}$ (recall the discussion at the end of Section 2). We proceed to establish this.

Fix any outcome ρ' of $\boldsymbol{\rho}' \leftarrow \mathcal{R}_p$ such that (i) the conclusion of Corollary 11 holds (i.e. $F \restriction \rho'$ is computed by a $(m/2, \{\mathsf{SYM}_{s^2}, \mathsf{THR}_{s^2}\} \circ \mathsf{DT}_{\log s})$-decision tree, which we call T), and (ii) the conclusion of Corollary 12 holds (i.e. $\mathrm{RW}_{m,k,r} \restriction \rho'$ contains a perfect copy of $\mathrm{RW}_{m,k,r'}$). (A random $\boldsymbol{\rho}' \leftarrow \mathcal{R}_p$ is such an outcome with probability at least $1 - \gamma_{\mathrm{SL}} - \gamma_{\mathrm{target}}$.) For ease of notation let us write RW' to denote $\mathrm{RW}_{m,k,r} \restriction \rho'$.

Fix any path π that reaches a leaf ℓ in T. (Note that a random choice of such a path corresponds to part (b) in the random draw of $\boldsymbol{\rho} \leftarrow \mathcal{R}$, recalling the discussion at the end of Section 2.) Since $|\pi| \leq m/2$, we have that the set

$$A_\ell := \{i \in [m] \colon \pi_{i,j,t} = * \text{ for all } j \in [k+1] \text{ and all } t\}$$

has cardinality at least $m - |\pi| \geq m/2$. In words, at least $m/2$ of the m many depth-2 subcircuits of RW' are completely "untouched" by π. For part (c) of the draw from $\mathcal{R}$, recall that the set $L(\ell)$ could be taken to be any good outcome of $\boldsymbol{L}(\ell)$, and that a random $\boldsymbol{L}(\ell) \subseteq_q [n]$ is good with probability at least $1 - 1/s$. By the same Chernoff bound argument as the one in Proposition 1, we have that

$$\begin{aligned}\Pr_{\boldsymbol{L} \subseteq_q [n]}\left[\exists\,(i,j) \in A_\ell \times [k+1] \colon \left|\{t \colon x_{i,j,t} \in \boldsymbol{L}(\ell)\}\right| < \frac{qr'}{2}\right] &\leq |A_\ell|(k+1)\exp\left(-\Omega(qr')\right) \\ &\ll \exp(-\Omega(n^{0.48})),\end{aligned}$$

recalling that $|A_\ell| \leq m$, $k = \Theta(\log n)$, $q \geq n^{-0.01}$ and $r' > n^{0.49}$. Since $1 - 1/s + 1 - \exp(-\Omega(n^{0.48})) > 1$, there must exist a good outcome $L(\ell)$ of $\boldsymbol{L}(\ell)$ such that for the

corresponding S_ℓ, every restriction ρ^{trim} fixing precisely the variables in S_ℓ is such that $\mathrm{RW}' \restriction \pi\rho^{\text{trim}}$ contains a perfect copy of

$$\mathrm{RW}_{m,k,r''}(x) = \bigoplus_{i \in A_\ell} \bigwedge_{j=1}^{k+1} \bigoplus_{t=1}^{r''} x_{i,j,t}, \qquad \text{where } r'' = \frac{qr'}{2} \gg 1.$$

Having $r'' \geq 1$ is crucial for us because, together with $|A_\ell| \geq m/2$, it means that $\mathrm{RW}_{m,k,r''}$ contains a perfect copy of $\mathrm{GIP}_{m/2,k+1}$ (i.e. by possibly restricting and renaming some variables of $\mathrm{RW}_{m,k,r''}$ and possibly negating the result, we obtain a function identical to $\mathrm{GIP}_{m/2,k+1}$).

▶ Remark. We refine the definition of $\mathcal{R}$ to require that in (c) it use an $L(\ell)$ as specified above at each leaf ℓ.

Summarizing, the above discussion establishes that $\mathrm{RW}_{m,k,r}$ "retains structure" with high probability under a random $\boldsymbol{\rho} \leftarrow \mathcal{R}$. The formal statement of this result (incorporating also Lemma 8) is as follows:

▶ **Lemma 13.** *Let F be any $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0{}_d$ circuit of size $s = n^{\tau \log n}$. The fair distribution $\mathcal{R}$ over restrictions $\boldsymbol{\rho} \in \{0,1,*\}^n$ from Lemma 8 satisfies the following: With probability $1 - \gamma_{\mathrm{SL}} - \gamma_{\text{target}}$ over a draw of $\boldsymbol{\rho} \leftarrow \mathcal{R}$, both of the following hold:*

(i) *$F \restriction \boldsymbol{\rho}$ belongs to $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AND}_k$;*

(ii) *$\mathrm{RW}_{m,k,r} \restriction \boldsymbol{\rho}$ contains a perfect copy of $\mathrm{GIP}_{m/2,k+1}$.*

4 Bounding the correlation between the approximator and target post-restriction

With Lemma 13 in hand it is a simple matter to finish the argument. Fix any outcome ρ of $\boldsymbol{\rho} \leftarrow \mathcal{R}$ such that $F \restriction \rho$ and $\mathrm{RW}_{m,k,r} \restriction \rho$ satisfy (i) and (ii) of Lemma 13. Applying either Fact 3 or Theorem 15 (depending on whether the top gate of F is SYM or THR) along with the lower bound of Theorem 14, we get that

$$\Pr_{\boldsymbol{x} \leftarrow \{0,1\}^n} \left[(F \restriction \rho)(\boldsymbol{x}) = (\mathrm{RW}_{m,k,r} \restriction \rho)(\boldsymbol{x}) \right] \leq \frac{1}{2} + \exp\left(-\Omega(m/4^k)\right) = 1/2 + \gamma_{\text{corr}}, \tag{5}$$

where

$$\gamma_{\text{corr}} = \exp\left(-\Omega(m/4^k)\right) = \exp\left(-\Omega(m^{0.999})\right) = \exp\left(-\Omega(n^{0.499})\right).$$

This gives ingredient (3) as described in Section 1.4. Recalling the discussion at the end of Section 1.4, Theorem 3 follows from Lemma 13 and (5).

References

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

2 Miklós Ajtai and Avi Wigderson. Deterministic simulation of probabilistic constant depth circuits. In *Proc. 26th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 11–19, 1985.

3 László Babai, Noam Nisan, and Márió Szegedy. Multiparty protocols, pseudorandom generators for logspace, and time-space trade-offs. *J. Comput. System Sci.*, 45(2):204–232, 1992. 21st Symposium on the Theory of Computing (STOC). doi:10.1016/0022-0000(92)90047-M.

4 Richard Beigel and Jun Tarui. On ACC. *Computational Complexity*, 4:350–366, 1994.

5 Manuel Blum and Silvio Micali. How to generate cryptographically strong sequences of pseudorandom bits. *SIAM J. Comput.*, 13(4):850–864, 1984. doi:10.1137/0213053.

6 Andrej Bogdanov. Pseudorandom generators for low degree polynomials. In *STOC'05: Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pages 21–30. ACM, New York, 2005. doi:10.1145/1060590.1060594.

7 Andrej Bogdanov and Emanuele Viola. Pseudorandom bits for polynomials. *SIAM J. Comput.*, 39(6):2464–2486, 2010. doi:10.1137/070712109.

8 Nader Bshouty. On learning multivariate polynomials under the uniform distribution. *Information Processing Letters*, 61(3):303–309, 1997.

9 Nader Bshouty and Yishay Mansour. Simple Learning Algorithms for Decision Trees and Multivariate Polynomials. *SIAM J. Comput.*, 31(6):1909–1925, 2002.

10 Shiteng Chen and Periklis A. Papakonstantinou. Depth reduction for composites. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science—FOCS 2016*, page to appear. IEEE, 2016.

11 Anindya De and Rocco Servedio. Efficient deterministic approximate counting for low-degree polynomial threshold functions. In *Proc. 46th Annual ACM Symposium on Theory of Computing (STOC)*, pages 832–841, 2014.

12 Ilias Diakonikolas, Daniel M. Kane, and Jelani Nelson. Bounded independence fools degree-2 threshold functions. In *Proc. 51st IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 11–20, 2010.

13 Ilias Diakonikolas, Homin Lee, Kevin Matulef, Krzysztof Onak, Ronitt Rubinfeld, Rocco Servedio, and Andrew Wan. Testing for concise representations. In *Proc. 48th Ann. Symposium on Computer Science (FOCS)*, pages 549–558, 2007.

14 Ilias Diakonikolas, Homin Lee, Krzysztof Matulef, Rocco Servedio, and Andrew Wan. Efficiently testing sparse GF(2) polynomials. *Algorithmica*, July 2010.

15 Ilias Diakonikolas, Ryan O'Donnell, Rocco Servedio, and Yi Wu. Hardness results for agnostically learning low-degree polynomial threshold functions. In *SODA*, pages 1590–1606, 2011.

16 A. Ehrenfeucht and M. Karpinski. The computational complexity of (xor,and)-counting problems. Technical report, preprint, 1989.

17 M. Goldmann. On the power of a threshold gate at the top. *Information Processing Letters*, 63(6):287–293, 1997.

18 D. Grigoriev, M. Karpinski, and M. Singer. Fast parallel algorithms for sparse multivariate polynomial interpolation over finite fields. *SIAM Journal on Computing*, 19(6):1059–1063, 1990.

19 Kristoffer Arnsfelt Hansen and Peter Bro Miltersen. Some meet-in-the-middle circuit lower bounds. In *Mathematical foundations of computer science 2004*, volume 3153 of *Lecture Notes in Comput. Sci.*, pages 334–345. Springer, Berlin, 2004. doi:10.1007/978-3-540-28629-5_24.

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

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

22 Johan Håstad and Mikael Goldmann. On the power of small-depth threshold circuits. *Comput. Complexity*, 1(2):113–129, 1991. doi:10.1007/BF01272517.

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

24 A. Kalai, A. Klivans, Y. Mansour, and R. Servedio. Agnostically learning halfspaces. *SIAM Journal on Computing*, 37(6):1777–1805, 2008.

25 Daniel Kane. A Structure Theorem for Poorly Anticoncentrated Gaussian Chaoses and Applications to the Study of Polynomial Threshold Functions. In *FOCS*, pages 91–100, 2012.

26 M. Karpinski. Boolean circuit complexity of algebraic interpolation problems. "ICSI technical report", 1989.

27 M. Karpinski and M. Luby. Approximating the Number of Zeros of a $GF[2]$ Polynomial. *Journal of Algorithms*, 14:280–287, 1993.

28 M. Krause and P. Pudlak. Computing boolean functions by polynomials and threshold circuits. *Computational Complexity*, 7(4):346–370, 1998.

29 Shachar Lovett. Unconditional pseudorandom generators for low-degree polynomials. *Theory Comput.*, 5:69–82, 2009. `doi:10.4086/toc.2009.v005a003`.

30 Shachar Lovett and Srikanth Srinivasan. Correlation bounds for poly-size AC^0 circuits with $n^{1-o(1)}$ symmetric gates. In *Approximation, randomization, and combinatorial optimization*, volume 6845 of *Lecture Notes in Comput. Sci.*, pages 640–651. Springer, Heidelberg, 2011. `doi:10.1007/978-3-642-22935-0_54`.

31 Michael Luby, Boban Veličković, and Avi Wigderson. Deterministic approximate counting of depth-2 circuits. In *Proceedings of the 2nd ISTCS*, pages 18–24, 1993.

32 Raghu Meka and David Zuckerman. Pseudorandom generators for polynomial threshold functions. In *STOC*, pages 427–436, 2010.

33 M. Minsky and S. Papert. *Perceptrons: an introduction to computational geometry*. MIT Press, Cambridge, MA, 1968.

34 Cody Murray and Ryan Williams. Circuit lower bounds for nondeterministic quasi-polytime: An easy witness lemma for NP and NQP. In *Proceedings of the 50th Annual Symposium on Theory of Computing (STOC)*, 2018.

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

36 Noam Nisan. The communication complexity of threshold gates. In *Combinatorics, Paul Erdős is eighty, Vol. 1*, Bolyai Soc. Math. Stud., pages 301–315. János Bolyai Math. Soc., Budapest, 1993.

37 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`.

38 V. V. Podolskii. Perceptrons of large weight. *Problems of Information Transmission*, 45(1):46–53, 2009.

39 Alexander Razborov and Avi Wigderson. $n^{\Omega(\log n)}$ lower bounds on the size of depth-3 threshold circuits with AND gates at the bottom. *Inform. Process. Lett.*, 45(6):303–307, 1993. `doi:10.1016/0020-0190(93)90041-7`.

40 R. Roth and G. Benedek. Interpolation and approximation of sparse multivariate polynomials over $GF(2)$. *SIAM J. Comput.*, 20(2):291–314, 1991.

41 Takayuki Sakai, Kazuhisa Seto, Suguru Tamaki, and Junichi Teruyama. Bounded depth circuits with weighted symmetric gates: Satisfiability, lower bounds and compression. In *41st International Symposium on Mathematical Foundations of Computer Science, MFCS 2016, August 22-26, 2016 - Kraków, Poland*, pages 82:1–82:16, 2016.

42 R. Schapire and L. Sellie. Learning sparse multivariate polynomials over a field with queries and counterexamples. *J. Comput. & Syst. Sci.*, 52(2):201–213, 1996.

43 Rocco A. Servedio and Li-Yang Tan. Improved pseudorandom generators from pseudorandom multi-switching lemmas, 2018. Manuscript. Available at `https://arxiv.org/abs/1801.03590`.

44 Adi Shamir. On the generation of cryptographically strong pseudorandom sequences. In *Automata, languages and programming (Akko, 1981)*, volume 115 of *Lecture Notes in Comput. Sci.*, pages 544–550. Springer, Berlin-New York, 1981.

45 Luca Trevisan and Tongke Xue. A derandomized switching lemma and an improved derandomization of AC^0 . In *Proceedings of the 28th IEEE Conference on Computational Complexity*, pages 242–247, 2013.

46 Salil P. Vadhan. Pseudorandomness. *Foundations and Trends in Theoretical Computer Science*, 7(1-3):1–336, 2012.

47 Emanuele Viola. Pseudorandom bits for constant-depth circuits with few arbitrary symmetric gates. *SIAM J. Comput.*, 36(5):1387–1403 (electronic), 2006/07. doi:10.1137/050640941.

48 Emanuele Viola. *On the power of small-depth computation.* Now Publishers Inc, 2009.

49 Emanuele Viola. The sum of d small-bias generators fools polynomials of degree d. *Comput. Complexity*, 18(2):209–217, 2009. doi:10.1007/s00037-009-0273-5.

50 Ryan Williams. Non-uniform ACC Circuit Lower Bounds. In *CCC 2011*, pages 115–125, 2011.

51 Andrew Yao. Theory and applications of trapdoor functions. In *23rd Annual Symposium on Foundations of Computer Science (Chicago, Ill., 1982)*, pages 80–91. IEEE, New York, 1982.

52 Andrew Yao. On ACC and threshold circuits. In *Proceedings of the Thirty-First Annual Symposium on Foundations of Computer Science*, pages 619–627, 1990.

A Preliminaries

We use bold font like $\boldsymbol{x}$, $\boldsymbol{\rho}$, etc. to denote random variables.

We write "size-S $\mathsf{AC}^0{}_d$" to denote the class of circuits of depth d consisting of at most S unbounded fan-in AND/OR gates with variables and negated variables as the inputs (we include these literals in the gate count).

Pseudorandomness. For $r < n$, we say that a distribution $\mathcal{D}$ over $\{0,1\}^n$ can be *sampled efficiently with r random bits* if (i) $\mathcal{D}$ is the uniform distribution over a multiset of size exactly 2^r of strings from $\{0,1\}^n$, and (ii) there is a deterministic algorithm $\mathrm{Gen}_{\mathcal{D}}$ which, given as input a uniform random r-bit string $\boldsymbol{x} \leftarrow \{0,1\}^r$, runs in time $\mathrm{poly}(n)$ and outputs a string drawn from $\mathcal{D}$.

For $\delta > 0$ and a class $\mathcal{C}$ of functions from $\{0,1\}^n$ to $\{0,1\}$, we say that a distribution $\mathcal{D}$ over $\{0,1\}^n$ *δ-fools $\mathcal{C}$ with seed length r* if (a) $\mathcal{D}$ can be sampled efficiently with r random bits via algorithm $\mathrm{Gen}_{\mathcal{D}}$, and (b) for every function $f \in \mathcal{C}$, we have

$$\left| \mathop{\mathbf{E}}_{\boldsymbol{s} \leftarrow \{0,1\}^r} [f(\mathrm{Gen}_{\mathcal{D}}(\boldsymbol{s}))] - \mathop{\mathbf{E}}_{\boldsymbol{x} \leftarrow \{0,1\}^n} [f(\boldsymbol{x})] \right| \leq \delta.$$

Equivalently, we say that $\mathrm{Gen}_{\mathcal{D}}$ is a *δ-PRG for $\mathcal{C}$ with seed length r*.

Restrictions. A *restriction* ρ of variables $x_1, \ldots, x_n$ is an element of $\{0,1,*\}^n$). 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 $\rho, \rho' \in \{0,1,*\}^n$, their *composition*, denoted $\rho\rho' \in \{0,1,*\}^n$, is the restriction defined by

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

We write $\mathcal{R}_p$ to denote the standard distribution over random restrictions with $*$-probability p, i.e. $\boldsymbol{\rho}$ drawn from $\mathcal{R}_p$ is a random string in $\{0,1,*\}$ obtained by independently setting each coordinate to $*$ with probability p and to each of $0,1$ with probability $\frac{1-p}{2}$.

A.1 Multiparty communication complexity

We recall a celebrated lower bound of Babai, Nisan, and Szegedy [3] on the multi-party "number on forehead" (NOF) communication complexity of the generalized inner product function:

▶ **Theorem 14** ([3]). *There is a partition of the $m \cdot (k+1)$ inputs of*

$$\mathrm{GIP}_{m,k+1}(x) := \bigoplus_{i=1}^{m} \bigwedge_{j=1}^{k+1} x_{i,j}$$

into $k+1$ blocks such that the following holds: Let P be a $(k+1)$-party randomized NOF communication protocol exchanging at most $\frac{1}{10}(m/4^{k+1} - \log(1/\gamma_{\mathrm{comm}}))$ bits of communication and computing a Boolean function f with error γ_{err} (meaning that on every input x the protocol outputs the correct value $f(x)$ with probability at least $1-\gamma_{\mathrm{err}}$). Then

$$\Pr_{\boldsymbol{x} \leftarrow \{0,1\}^{m(k+1)}} \left[f(\boldsymbol{x}) = \mathrm{GIP}_{m,k+1}(\boldsymbol{x}) \right] \leq \frac{1}{2} + \gamma_{\mathrm{err}} + \gamma_{\mathrm{comm}}.$$

The connection between $\mathsf{SYM} \circ \mathsf{AND}_k$ circuits and $(k+1)$-party communication complexity is due to the following simple but influential observation of Håstad and Goldmann:

▶ **Fact 3** ([22]). *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function computed by a size-s $\mathsf{SYM} \circ \mathsf{AND}_k$ circuit. Then for any partition of the n inputs of f into $k+1$ blocks, there is a deterministic NOF $(k+1)$-party communication protocol that computes f using $O(k \log s)$ bits of communication.*

For $\mathsf{THR} \circ \mathsf{AC}^0$ circuits we use an analogous result from [36] on the $(k+1)$-party randomized γ-error communication complexity of $\mathsf{THR} \circ \mathsf{AND}_k$ circuits:

▶ **Theorem 15** ([36]). *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function computed by a $\mathsf{THR} \circ \mathsf{AND}_k$ circuit. Then for any partition of the n inputs of f into $k+1$ blocks, there is a randomized NOF $(k+1)$-party communication protocol that computes f with error γ_{err} using $O(k^3 \log n \log(n/\gamma_{\mathrm{err}}))$ bits of communication.*

B Applying the [37] paradigm to obtain pseudorandom generators from correlation bounds

A function f is said to be (s,τ)-*hard for a circuit class* $\mathcal{C}$ if every circuit $C \in \mathcal{C}$ of size at most s has $\Pr_{\boldsymbol{x}}[f(\boldsymbol{x}) = C(\boldsymbol{x})] \leq \frac{1}{2} + \tau$, where $\boldsymbol{x}$ is a uniform random input string. If this holds then we say that f gives a *correlation bound* of τ against $\mathcal{C}$-circuits of size s.

Given a quadruple (m,r,ℓ,s) of non-negative integers, a family $\mathcal{F} = \{T_1,\ldots,T_s\}$ of r-element subsets of $[m]$ is said to be an (m,r,ℓ,s)-*design* if for any two distinct subsets $T_i, T_j \in \mathcal{F}$ we have $|T_i \cap T_j| \leq \ell$.

An ANY_t gate is a gate that takes in t inputs and computes an arbitrary function from $\{0,1\}^t$ to $\{0,1\}$.

We recall the Nisan-Wigderson [37] translation from correlation bounds to PRGs:

▶ **Theorem 16** (The Nisan-Wigderson generator). *Fix a circuit class $\mathcal{C}$ and let $m, r, \ell, s \in \mathbb{N}$ be positive parameters with $m \geq r \geq \ell$. Given an explicit $f : \{0,1\}^r \to \{0,1\}$ that is $(s \cdot 2^\ell, \varepsilon/s)$-hard for $\mathcal{C} \circ \mathsf{ANY}_{\log \ell}$ and an explicit (m, r, ℓ, s)-design, there is an explicit PRG $G : \{0,1\}^m \to \{0,1\}^s$ that ε-fools size-s circuits in $\mathcal{C}$. (Hence for $s \geq n$, by taking the first n output bits of G there is an explicit PRG mapping $\{0,1\}^m$ to $\{0,1\}^n$ that ε-fools size-s n-variable circuits in $\mathcal{C}$.)*

The existence of explicit designs is well known, in particular we recall the following:

▶ **Lemma 17** (Problem 3.2 of [46]). *There is a deterministic algorithm which, for any $r, s \in \mathbb{N}$, runs in time $\mathrm{poly}(m, s)$ and outputs an explicit (m, r, ℓ, s)-design with $m = O(r^2/s)$ and $\ell \leq \log s$.*

A PRG from the [47] correlation bound. Viola [47] gives an explicit function $f : \{0,1\}^r \to \{0,1\}$ and shows that for every constant d there is a constant c_d such that f is $(r^{c_d \log r}, r^{-c_d \log r})$-hard for $\mathsf{SYM} \circ \mathsf{AC}^0{}_d$. Fix any d. Given values for ε, s let us set the parameters

$$\ell = \log s, \qquad r = 2^{10 \cdot \sqrt{\frac{1}{c_d} \log(s/\varepsilon)}}.$$

It is straightforward to verify that $s \cdot 2^\ell \leq r^{c_d \log r}$ and $\varepsilon/s \geq r^{-c_d \log r}$. By Lemma 17 there is an explicit (m, r, ℓ, s)-design with $m = O(r^2/\ell) = 2^{O(\sqrt{\frac{1}{c_d} \log(s/\varepsilon)})}$, so applying the Nisan-Wigderson generator, we get that for $s \geq n$, there is an explicit PRG $G : \{0,1\}^m \to \{0,1\}^n$ with seed length $m = 2^{O(\sqrt{\log(s/\varepsilon)})}$ that ε-fools n-variable size-s circuits in $\mathsf{AC}^0{}_d$.

A PRG from the [30] correlation bound. Lovett and Srinivasan [30] give an explicit function $f : \{0,1\}^r \to \{0,1\}$ such that for every constant d there is a constant c_d such that f is $(r^{c_d \log \log r}, \exp(-r^{1-o(1)})$-hard for $\mathsf{SYM} \circ \mathsf{AC}^0{}_d$. We proceed as above but now choosing

$$\ell = \log s, \qquad r = 2^{\frac{10}{c_d} \cdot \frac{\log s}{\log \log s}} + (\log(s/\varepsilon))^{1+o(1)}.$$

It is straightforward to verify that $s \cdot 2^\ell \leq r^{c_d \log \log r}$ and $\varepsilon/s \geq \exp(-r^{1-o(1)})$. By Lemma 17 there is an explicit (m, r, ℓ, s)-design with $m = O(r^2/\ell) = 2^{O(\log s/\log \log s)} \cdot (\log(1/\varepsilon))^{2+o(1)}$, so applying the Nisan-Wigderson generator, we get that for $s \geq n$, there is an explicit PRG $G : \{0,1\}^m \to \{0,1\}^n$ with seed length $m = 2^{O(\log s/\log \log s)} + (\log(1/\varepsilon))^{2+o(1)}$ that ε-fools n-variable size-s circuits in $\mathsf{SYM} \circ \mathsf{AC}^0{}_d$.

For $\mathsf{THR} \circ \mathsf{AC}^0{}_d$, [30] show that the same function f is $(r^{c_d \log \log r}, \exp(-r^{1/2-o(1)}))$-hard for $\mathsf{THR} \circ \mathsf{AC}^0{}_d$; a similar analysis to the above gives an explicit PRG $G : \{0,1\}^m \to \{0,1\}^n$ with seed length $m = 2^{O(\log s/\log \log s)} + (\log(1/\varepsilon))^{4+o(1)}$ that ε-fools n-variable size-s circuits in $\mathsf{THR} \circ \mathsf{AC}^0{}_d$.

A PRG from our Theorem 3: Proof of Theorem 2. Theorem 3 gives an explicit $f : \{0,1\}^r \to \{0,1\}$ and $\tau > 0$ such that for all d, f is $(r^{\tau \log r}, \exp(-r^{0.499}))$-hard for $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0{}_d$. This time we choose

$$\ell = \log s, \qquad r = 2^{10 \cdot \sqrt{\frac{2}{\tau} \log s}} + (\log(s/\varepsilon))^{2.005}.$$

We have $s \cdot 2^\ell \leq r^{\tau \log r}$ and $\varepsilon/s \geq \exp(-r^{0.499})$, so we get that for $s \geq n$, there is an explicit PRG $G : \{0,1\}^m \to \{0,1\}^n$ with seed length $m = O(r^2/\ell) = 2^{O(\sqrt{\log s})} + (\log(1/\varepsilon))^{4.01}$ that ε-fools n-variable size-s circuits in $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0{}_d$.

A PRG from our Theorem 4: Proof of Corollary 5. Finally, Theorem 4 gives an explicit $f : \{0,1\}^r \to \{0,1\}$ and $\tau > 0$ such that for all d, f is $(r^{\tau \log r}, \exp(-r^{0.499}))$-hard for the class of depth-d circuits over $\{0,1\}^r$ that contain $r^{0.249}$ many SYM or THR gates. We choose ℓ, r as above, so similar to the above, we get that there is an explicit PRG $G : \{0,1\}^m \to \{0,1\}^n$ with seed length $m = O(r^2/\ell) = 2^{O(\sqrt{\log s})} + (\log(1/\varepsilon))^{4.01}$ that ε-fools n-variable size-s depth-d circuits with at most $2^{c\sqrt{\log s}}$ many SYM or THR gates.

C Proof of Theorem 4: Handling multiple SYM and THR gates

We prove Theorem 4 via a slight variant of Theorem 3 and an argument from [30] (a related argument appears in a somewhat different form in [47]). The variant of Theorem 3, stated as Theorem 20 below, is proved by combining ingredients (1), (2) and (3) as in Section 1.4, but now with the aim of proving a correlation bound against $\mathsf{ANY}_u \circ \{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0{}_d$ circuits rather than $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0{}_d$ circuits (where here and throughout this appendix we take $u := n^{0.249}$). As we describe at the end of this section, once this correlation bound against $\mathsf{ANY}_u \circ \{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0{}_d$ is in place, the extension to circuits with $n^{0.249}$ many SYM or THR gates directly follows using an argument from [30].

In more detail we have: (throughout the following the values of m, k, r are as they were before)

▶ **Lemma 18** (Lemma 8 analogue). *Fix $u := n^{0.249}$ and let F be an $\mathsf{ANY}_u \circ \{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0{}_d$ circuit where each of the u $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0{}_d$ subcircuits of F has size at most $s = n^{\tau \log n}$. There is a fair distribution $\mathcal{R}$ over restrictions $\boldsymbol{\rho} \in \{0,1,*\}^n$ such that the following holds: With probability $1 - \gamma_{\mathrm{SL}} = 1 - \exp(-\Omega_d(\sqrt{n/\log n}))$ over the draw of $\boldsymbol{\rho} \leftarrow \mathcal{R}$, it is the case that $F \restriction \boldsymbol{\rho}$ belongs to the class $\mathcal{F}_{simple,\,u} := \mathsf{ANY}_u \circ \{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AND}_k$.*

The proof is almost identical to that of Lemma 8, with $\mathsf{ANY}_u \cdot \{\mathsf{SYM}, \mathsf{THR}\}$ taking the place of $\{\mathsf{SYM}, \mathsf{THR}\}$ throughout the argument. Now in Corollary 10 the gate G corresponds to $\mathsf{ANY}_u \circ \{\mathsf{SYM}, \mathsf{THR}\}$ (rather than to just $\{\mathsf{SYM}, \mathsf{THR}\}$ as earlier) and the total circuit size of F is us rather than s (leading to $us \cdot 2^{-t}$ rather than $s \cdot 2^{-t}$ on the RHS of the Corollary 10 bound), but this is swallowed up by the slack in the inequalities leading to (1).

▶ **Lemma 19** (Lemma 13 analogue). *Fix $u := n^{0.249}$ and let F be an $\mathsf{ANY}_u \circ \{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0{}_d$ circuit where each $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0{}_d$ subcircuit has size at most $s = n^{\tau \log n}$. The fair distribution $\mathcal{R}$ over restrictions $\boldsymbol{\rho} \in \{0,1,*\}^n$ from Lemma 18 satisfies the following: With probability $1 - \gamma_{\mathrm{SL}} - \gamma_{\mathrm{target}}$ over a draw of $\boldsymbol{\rho} \leftarrow \mathcal{R}$, both of the following hold:*

- **(i)** *$F \restriction \boldsymbol{\rho}$ belongs to $\mathcal{F}_{simple,\,u} = \mathsf{ANY}_u \circ \{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AND}_k$; and*
- **(ii)** *$\mathrm{RW}_{m,k,r} \restriction \boldsymbol{\rho}$ contains a perfect copy of $\mathrm{GIP}_{m/2,k+1}$.*

The proof of Lemma 19 is unchanged from Section 3.

▶ **Theorem 20** (Theorem 3 analogue). *Fix $u := n^{0.249}$. There is an absolute constant $\tau > 0$ and an explicit $\mathrm{poly}(n)$-time computable function $H : \{0,1\}^n \to \{0,1\}$ with the following property: for any constant d, for n sufficiently large, for F an $\mathsf{ANY}_u \circ \{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0{}_d$ circuit where each $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0{}_d$ subcircuit has size at most $s = n^{\tau \log n}$, we have*

$$\Pr_{\boldsymbol{x} \leftarrow \{0,1\}^n}[F(\boldsymbol{x}) = \mathrm{RW}_{m,k,r}(\boldsymbol{x})] \le \frac{1}{2} + \exp(-\Omega(n^{0.249})).$$

The proof, using Lemmas 18 and 19, is virtually identical to the proof of Theorem 3 using Lemmas 8 and 13. The only difference is that we use the obvious extensions of Fact 3 and Theorem 15 to $\mathsf{ANY}_u \cdot \mathsf{SYM} \circ \mathsf{AND}_k$ circuits and $\mathsf{ANY}_u \cdot \mathsf{THR} \circ \mathsf{AND}_k$ circuits respectively; these extensions are stated for completeness below.

▶ **Fact 4** (Fact 3 analogue). *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function computed by a size-s $\mathsf{ANY}_u \circ \mathsf{SYM} \circ \mathsf{AND}_k$ circuit. Then for any partition of the n inputs of f into $k+1$ blocks, there is a deterministic NOF $(k+1)$-party communication protocol that computes f using $u \cdot O(k \log s)$ bits of communication.*

▶ **Theorem 21** (Theorem 15 analogue). *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function computed by a $\mathsf{ANY}_u \circ \mathsf{THR} \circ \mathsf{AND}_k$ circuit. Then for any partition of the n inputs of f into $k+1$ blocks, there is a randomized NOF $(k+1)$-party communication protocol that computes f with error γ_{err} using $u \cdot O(k^3 \log n \log(n/\gamma_{\mathrm{err}}))$ bits of communication.*

We note that the extra factors of u in Fact 4 and in Theorem 21, which are not present in the analogous Fact 3 and Theorem 15 respectively, are responsible for the quantitatively weaker correlation bound in Theorem 4 as opposed to Theorem 3. The $n^{0.249}$ bound on the number of SYM or THR gates which Theorem 4 can handle can be traded off against the correlation bound in that theorem; we leave the details to the interested reader.

Finally, the correlation bound Theorem 4 follows from Theorem 20 exactly as Theorem 6 of [30] follows from Lemma 3 of that paper.