Electronic Colloquium on Computational Complexity, Revision 2 of Report No. 183 (2018)

Near-Optimal Pseudorandom Generators for Constant-Depth Read-Once Formulas

Dean Doron*
Department of Computer Science
University of Texas at Austin
deandoron@utexas.edu

Pooya Hatami†
Department of Computer Science
University of Texas at Austin
pooyahat@gmail.com

William M. Hoza‡
Department of Computer Science
University of Texas at Austin
whoza@utexas.edu

Abstract

We give an explicit pseudorandom generator (PRG) for read-once $\mathbf{AC}^0$, i.e., constant-depth read-once formulas over the basis $\{\wedge, \vee, \neg\}$ with unbounded fan-in. The seed length of our PRG is $\widetilde{O}(\log(n/\varepsilon))$. Previously, PRGs with near-optimal seed length were known only for the depth-2 case [GMR+12]. For a constant depth $d > 2$, the best prior PRG is a recent construction by Forbes and Kelley with seed length $\widetilde{O}(\log^2 n + \log n \log(1/\varepsilon))$ for the more general model of constant-width read-once branching programs with arbitrary variable order [FK18]. Looking beyond read-once $\mathbf{AC}^0$, we also show that our PRG fools read-once $\mathbf{AC}^0[\oplus]$ with seed length $\widetilde{O}(t + \log(n/\varepsilon))$, where t is the number of parity gates in the formula.

Our construction follows Ajtai and Wigderson's approach of iterated pseudorandom restrictions [AW89]. We assume by recursion that we already have a PRG for depth-d $\mathbf{AC}^0$ formulas. To fool depth-$(d+1)$ $\mathbf{AC}^0$ formulas, we use the given PRG, combined with a small-bias distribution and almost k-wise independence, to sample a pseudorandom restriction. The analysis of Forbes and Kelley [FK18] shows that our restriction approximately preserves the expectation of the formula. The crux of our work is showing that after $\mathrm{poly}(\log\log n)$ independent applications of our pseudorandom restriction, the formula simplifies in the sense that every gate other than the output has only $\mathrm{polylog}\, n$ remaining children. Finally, as the last step, we use a recent PRG by Meka, Reingold, and Tal [MRT19] to fool this simpler formula.

*Supported by NSF Grant CCF-1705028.
†Supported by a Simons Investigator Award (#409864, David Zuckerman).
‡Supported by the NSF GRFP under Grant DGE-1610403 and by a Harrington Fellowship from UT Austin.

ISSN 1433-8092

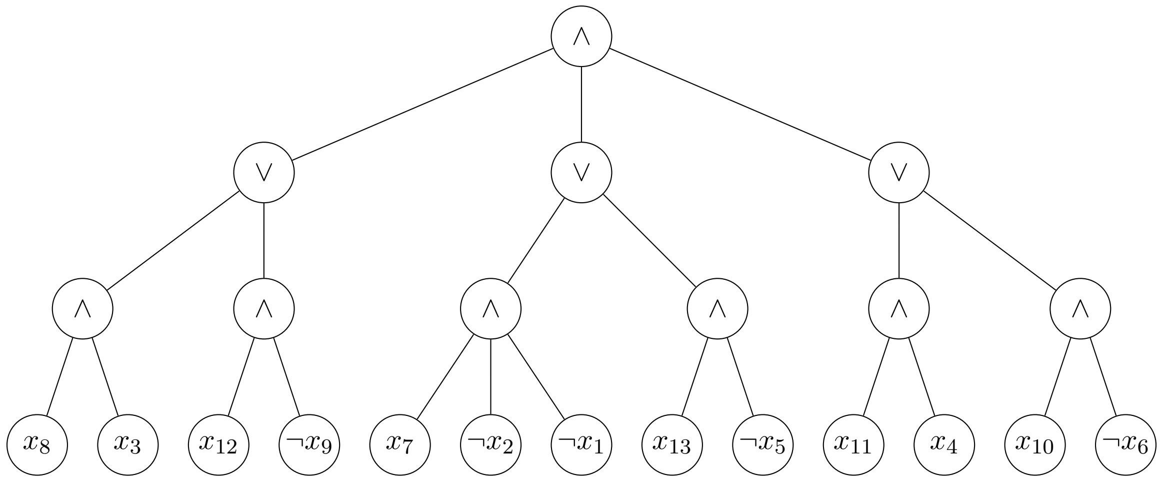

Figure 1: A depth-3 read-once $\mathbf{AC}^0$ formula on $n = 13$ bits.

1 Introduction

In complexity theory and algorithm design, randomness is a valuable yet scarce resource. A powerful, black-box method for reducing the randomness used by a computationally bounded process is to construct a *pseudorandom generator* (PRG). A PRG for a class of tests $\mathcal{C}$ is an algorithm that stretches a short truly random seed to a long n-bit string that "fools" $\mathcal{C}$, i.e., any test $f \in \mathcal{C}$ behaves the same on the output of the PRG as it does on a truly random string, up to some error ε.

Ideally, one would like to construct explicit unconditional PRGs with short seed length that fool powerful classes such as general polynomial-time algorithms. Unfortunately, constructing such general-purpose PRGs requires proving circuit lower bounds that seem to be far beyond the reach of state of the art techniques.

On the bright side, there has been a lot of success designing PRGs for more restricted classes. The two most intensely studied classes are read-once small-space algorithms and constant-depth circuits. In this work, we study *constant-depth read-once formulas* with unbounded fan-in over the basis $\{\wedge, \vee, \neg\}$ (Figure 1). This class is the read-once version of $\mathbf{AC}^0$. We construct an explicit PRG for this class with seed length $\widetilde{O}(\log(n/\varepsilon))$, which is optimal up to $\log\log$ factors.

Theorem 1.1. *For any positive integers n, d and for any $\varepsilon > 0$, there is an explicit ε-PRG for depth-d read-once $\mathbf{AC}^0$ formulas over n variables with seed length*

$$\log(n/\varepsilon) \cdot O(d \log\log(n/\varepsilon))^{2d+2}.$$

1.1 Motivation and related work

Derandomizing small-space algorithms. We are motivated by the $\mathbf{L}$ vs. $\mathbf{BPL}$ problem – namely whether every bounded-error probabilistic algorithm can be fully derandomized with only a constant factor space blowup. The way a log-space algorithm acts on its random bits can be modeled by a polynomial-width *read-once branching program* (ROBP). A natural approach to the $\mathbf{L}$ vs. $\mathbf{BPL}$ problem is thus coming up with a PRG for such ROBPs with seed length $O(\log n)$. Seminal work of Nisan gave a PRG with seed length $O(\log^2 n)$ for this model [Nis92]. To this day, no better PRG is known even for ROBPs where the width is a large *constant*, though better generators are known in special cases [ŠŽ11, De11, KNP11, Ste12, BRRY14, GMR+12, BDVY13, CHHL18, MRT19].

Surprisingly, the study of fooling constant-width ROBPs has so far been closely entangled with the study of fooling read-once $\mathbf{AC}^0$. A depth-d read-once $\mathbf{AC}^0$ formula can be computed by a width-$(d+1)$ ROBP, possibly after reordering the inputs [CSV15]. In the other direction, Gopalan et al. constructed a near-optimal PRG for read-once CNFs, and then used that PRG to construct a near-optimal hitting set generator for width-3 ROBPs [GMR+12]. Very recently, following the paradigm of Gopalan et al. [GMR+12], Meka, Reingold, and Tal gave a PRG for general width-3 ROBPs with near-optimal seed length when ε is constant [MRT19].

Meanwhile, for any constant d, Chen, Steinke and Vadhan constructed a PRG for depth-d read-once $\mathbf{AC}^0$ formulas with seed length $\widetilde{O}(\log^{d+1} n)$ [CSV15].[1] They obtained this PRG by proving new Fourier tail bounds for such formulas. Subsequently, Chattopadhyay et al. proved similar tail bounds for the stronger class of general width-$(d+1)$ ROBPs with arbitrarily ordered inputs; they used these tail bounds to construct a PRG with similar seed length for that model [CHRT18].

In a recent breakthrough, Forbes and Kelley gave an elegant construction of a PRG for ROBPs with arbitrarily ordered inputs [FK18]. In the polynomial-width case, their PRG has seed length $O(\log^3 n)$. For width-$(d+1)$ ROBPs when d is small, their PRG has seed length $\widetilde{O}(d \log^2 n)$; prior to the present work, this was also the best PRG for read-once $\mathbf{AC}^0$. Note that Theorem 1.1 improves on the Forbes-Kelley PRG [FK18] even for non-constant d, e.g., if $d = 0.2 \log\log n / \log\log\log n$ and $\varepsilon = 1/\operatorname{poly}(n)$.

Given the recent trend of connections between PRGs for ROBPs and PRGs for read-once $\mathbf{AC}^0$, we hope that our result will serve as a stepping stone toward optimal PRGs for general constant-width ROBPs.

Fooling general constant-depth circuits. Ajtai and Wigderson were the first to consider the problem of fooling general $\mathbf{AC}^0$ circuits, and in their pioneering work they achieved seed length $O(n^\gamma)$ for any constant $\gamma > 0$ [AW89]. A long line of research has worked on improving this seed length [Nis91, LN90, LVW93, Baz09, Raz09, Bra09, DETT10, GMR13, TX13, Tal17, HS16, ST18]. Today, for constant error, the best PRG for depth-d $\mathbf{AC}^0$ circuits known, by Tal, has seed length $\widetilde{O}(\log^{d+2} n)$ [Tal17]. When ε is small, the best PRG is a very recent construction by Servedio and Tan [ST18], which achieves seed length $O(\log^{d+C} n \log(1/\varepsilon))$ for some unspecified absolute constant C.

Fooling more general read-once formulas. Bogdanov, Papakonstantinou, and Wan gave the first PRG for *unbounded-depth* read-once formulas [BPW11]. Their PRG has seed length $(1-\Omega(1))n$ and fools read-once formulas over the $\{\wedge, \vee, \neg\}$ basis with unbounded fan-in. Their PRG also fools formulas over an arbitrary basis with fan-in $O(n/\log n)$. For the case that the basis is $\{\wedge, \vee, \neg\}$ and the fan-in is 2, Impagliazzo, Meka, and Zuckerman gave an improved PRG for unbounded-depth read-once formulas with seed length $O(n^{0.2342})$ [IMZ12]. The recent PRG by Forbes and Kelley [FK18] with seed length $O(\log^3 n)$ fools unbounded-depth read-once formulas either in the case that the basis is $\{\wedge, \vee, \neg\}$ and the fan-in is unbounded or the case of arbitrary basis and constant fan-in.

In another direction, Gavinsky, Lovett, and Srinivasan gave a PRG for constant-depth read-once formulas over the basis $\{\wedge, \vee, \neg, \mathrm{MOD}_m\}$, i.e., read-once $\mathbf{ACC}^0$ [GLS12]. When the modulus m and the error ε are constant, their PRG has seed length $2^{O(d^2)} \cdot \log^{O(d)} n$; this result is also subsumed

[1] Note that Nisan's generator [Nis92] is not guaranteed to fool read-once $\mathbf{AC}^0$ formulas because of the issue of variable ordering [BPW11].

by the recent work of Forbes and Kelley [FK18]. As a reminder, in the present work, we focus on constant-depth read-once formulas over the $\{\wedge, \vee, \neg\}$ basis with unbounded fan-in.

Fooling read-k depth-2 formulas. De et al. gave a PRG for read-once CNFs with seed length $O(\log n \log(1/\varepsilon))$ [DETT10]; this result can also be deduced from earlier work by Chari, Rohatgi, and Srinivasan [CRS00]. As mentioned previously, Gopalan et al. gave a PRG for read-once CNFs with seed length $\widetilde{O}(\log(n/\varepsilon))$ [GMR+12]. Meanwhile, Klivans, Lee, and Wan constructed a PRG that fools read-k CNFs even for small $k > 1$ [KLW10]. Building on their work, Servedio and Tan recently gave an improved PRG for read-k CNFs [ST19]; if the size of the CNF is $\mathrm{poly}(n)$, their PRG has seed length $\log n \cdot \mathrm{poly}(k, \log(1/\varepsilon))$.

1.2 Overview of our construction and analysis

1.2.1 The Ajtai-Wigderson approach [AW89]

Our PRG follows the paradigm pioneered by Ajtai and Wigderson [AW89] and further developed by Gopalan et al. [GMR+12]. We begin by briefly explaining this general approach for constructing PRGs. Ultimately, to fool a test f, we want to pseudorandomly assign values to its inputs in such a way that f accepts or rejects with approximately the same probability as it would under a truly random input. As a first step, we pseudorandomly choose a *partial* assignment to f. Equivalently, we pseudorandomly choose a *restriction* $X \in \{0, 1, \star\}^n$, where $X_i = \star$ indicates that the variable X_i is still unset.

We need our pseudorandom distribution over restrictions to satisfy two key properties. The first property is that the restriction should approximately *preserve the expectation* of the function, i.e., in expectation over X, the restricted function $f|_X$ should have approximately the same bias as f itself. This feature ensures that after sampling the pseudorandom restriction X, our remaining task is simply to fool the restricted function $f|_X$.

The second property is that the restriction should *simplify* f, i.e., with high probability[2] over the pseudorandom restriction X, the restricted function $f|_X$ should in some sense be simpler than f itself. The purpose of this feature is that simplifying f should make it easier to fool, perhaps using a PRG from prior work. We shall now give a brief exposition of how we achieve these two properties in our work.

1.2.2 Preserving the expectation using the work of Forbes and Kelley [FK18]

Building on several prior works [RSV13, HLV18, CHRT18], Forbes and Kelley constructed a very simple pseudorandom distribution over restrictions that approximately preserves the expectation of any constant-width ROBP [FK18], hence any read-once $\mathbf{AC}^0$ formula. In the Forbes-Kelley distribution, the locations of the $\star$-s are chosen almost k-wise independently, and the non-$\star$ coordinates are filled in using a small-bias space. Each coordinate is $\star$ with probability roughly $\frac{1}{2}$, and the distribution can be sampled using $\widetilde{O}(\log(n/\varepsilon))$ truly random bits.

In our setting, we will design our restriction in such a way that the distribution of $\star$ locations is almost k-wise independent and the distribution of bits in the non-$\star$ coordinates has small bias, in addition to other properties we also need. That way, to argue that the expectation of the formula is preserved under our pseudorandom restriction, we can simply appeal to the Forbes-Kelley result [FK18].

[2] In principle, it would actually suffice for f to merely simplify *in expectation* over X.

1.2.3 Simplifying the formula *given* a PRG

The remaining challenge is to ensure that our pseudorandom restriction *simplifies* $\mathbf{AC}^0$ formulas. In the work of Forbes and Kelley [FK18], the measure of complexity was simply the number of remaining unset variables. That is, Forbes and Kelley argued that after applying $O(\log n)$ independent pseudorandom restrictions, with high probability, all variables are set, and hence there is nothing left to fool [FK18].[3] This gives them an overall seed length of $\widetilde{O}(\log(n/\varepsilon)\log n)$.

In this work, to achieve seed length $\widetilde{O}(\log(n/\varepsilon))$, we use a more sophisticated pseudorandom restriction and subtler measures of complexity. That way, we can argue that after applying just $\mathrm{poly}(\log\log(n/\varepsilon))$ independent restrictions, the formula has simplified enough that it can be fooled by a prior PRG.

Several "pseudorandom switching lemmas" are already known for $\mathbf{AC}^0$ [AW89, TX13, GW14, ST18], but we were not able to use these lemmas for our result. Instead, the starting point for our approach to simplification is the work of Chen, Steinke, and Vadhan [CSV15]. Chen et al. analyzed the effect of *truly* random restrictions on read-once $\mathbf{AC}^0$ formulas [CSV15]. They showed that with high probability, a truly random restriction dramatically simplifies the formula in the sense that every node in the restricted formula has very few remaining children[4] [CSV15]. Chen et al. mentioned that they would have liked to show that the same is true under pseudorandom restrictions – this would have improved the parameters of their main result – but they were not able to prove such a statement [CSV15].

A key insight in our work is that roughly speaking, the predicate that some node is still alive after a random restriction X can be computed by *another read-once* $\mathbf{AC}^0$ *formula* whose inputs are the bits encoding X. Therefore, to pseudorandomly sample a restriction X that kills off each node with approximately the right probability, it suffices to select the bits encoding X using a PRG for read-once $\mathbf{AC}^0$. (Gavinsky, Lovett, and Srinivasan used a similar idea to fool read-once $\mathbf{ACC}^0$ [GLS12].)

1.2.4 Obtaining the necessary PRG through recursion

It may strike the reader that we have reached a "chicken or egg" problem: we can simplify formulas *given* a PRG for read-once $\mathbf{AC}^0$, but the whole reason we are interested in simplifying formulas is to *design* an improved PRG for read-once $\mathbf{AC}^0$! We resolve this difficulty by *recursing* on the depth of the formula we wish to fool. That is, we assume we already have a PRG G_d that fools depth-d read-once $\mathbf{AC}^0$ formulas, and we use G_d to sample pseudorandom restrictions that simplify depth-$(d+1)$ read-once $\mathbf{AC}^0$ formulas. (This is similar to the approach of Gavinsky et al. [GLS12].) Making this idea work requires overcoming several technical challenges.

In more detail, consider a collection of nodes $\{\phi_1, \dots, \phi_k\}$ that form subformulas of depth $d' \le d-1$. Roughly speaking, we show how to test the predicate that they are all still alive by a formula T of depth $d'+1 \le d$.[5] The recursive generator G_d fools T, so under our pseudorandom restriction, the probability that $\phi_1, \dots, \phi_k$ all remain alive is roughly what it would be under a truly random restriction.

Unfortunately, to ensure that the Forbes-Kelley analysis applies to our scenario, we are forced to design our pseudorandom restriction so that each coordinate is $\star$ with constant probability. The pseudorandom restriction has a similar effect as a truly random restriction with the same

[3] Actually, to get the best dependence on ε, Forbes and Kelley stop applying restrictions once the number of remaining variables drops below $O(\log n)$.

[4] A technicality is that this is only true "up to sandwiching."

[5] See Claim 5.5 for the precise statement.

$\star$-probability, but that is not good enough. The analysis of truly random restrictions by Chen et al. only applies to the case that the $\star$-probability is $1/\operatorname{polylog}(n/\varepsilon)$ [CSV15].

Roughly speaking, we overcome this difficulty using a kind of hybrid argument. A truly random restriction with $\star$-probability $1/\operatorname{polylog}(n/\varepsilon)$ is equivalent to the composition of t independent truly random restrictions, each with constant $\star$-probability, where $t = O(\log\log(n/\varepsilon))$. We show that for the purpose of simplification, a composition of t independent copies of our pseudorandom restriction is almost as good. Each individual step of this hybrid argument relies on the fact that G_d fools a formula closely related to the formula T mentioned earlier.

By applying an argument due to Gopalan et al. [GMR+12], we relate the condition that a collection of gates all remain alive to the number of remaining children of each node. Altogether, these arguments show that after applying $\operatorname{poly}(\log\log(n/\varepsilon))$ independent copies of our pseudorandom restriction, every gate *other than the root* has at most $\operatorname{polylog}(n)$ remaining children.[6] (We are not able to establish such a bound for the root gate, because its children form subformulas of depth $d' = d$.) Fortunately, this condition is strong enough that the restricted formula is fooled by a recent PRG by Meka, Reingold, and Tal [MRT19]. We use the MRT PRG [MRT19] as the last step in our construction.

1.3 Extension to read-once $\mathbf{AC}^0[\oplus]$ with a few parity gates

$\mathbf{AC}^0$ is admittedly a fairly weak circuit class. The parity function is the most famous example of a function that cannot be computed in $\mathbf{AC}^0$ (e.g., [FSS84, Has86]). Having shown how to fool read-once $\mathbf{AC}^0$, the natural next problem is to fool read-once $\mathbf{AC}^0[\oplus]$, i.e., constant-depth read-once formulas over the basis $\{\oplus, \wedge, \vee, \neg\}$ with unbounded fan-in. Read-once $\mathbf{AC}^0[\oplus]$ can still be simulated by constant-width ROBPs (possibly after reordering the inputs), so fooling read-once $\mathbf{AC}^0[\oplus]$ would be another step on the long road to derandomizing $\mathbf{BPL}$. The best prior PRG for this model is once again Forbes and Kelley's PRG with seed length $\widetilde{O}(\log^2 n + \log n \log(1/\varepsilon))$ [FK18].

Fooling general (not necessarily read-once) $\mathbf{AC}^0[\oplus]$ circuits is a notoriously difficult problem in unconditional pseudorandomness. Currently, the best seed length is only slightly less than n [FSUV13].

There has been more success fooling $\mathbf{AC}^0[\oplus]$ circuits under the assumption that the circuit only has a few parity gates [Vio07, CH05, LS11]. In the same spirit, we show that our PRG for read-once $\mathbf{AC}^0$ formulas also fools read-once $\mathbf{AC}^0[\oplus]$ formulas with a bounded number of parity gates. We achieve seed length $\widetilde{O}(t + \log(n/\varepsilon))$, where t is the number of parity gates:

Theorem 1.2. *For any positive integers n, d, t and for any $\varepsilon > 0$, there is an explicit ε-PRG for depth-d read-once $\mathbf{AC}^0[\oplus]$ formulas with at most t parity gates with seed length*

$$(td + \log(n/\varepsilon)) \cdot O(d \log\log(n/\varepsilon) + d\log(td))^{2d+2}.$$

At a very high level, this extension to $\mathbf{AC}^0[\oplus]$ is possible because the MRT PRG [MRT19] was already designed for parities of small ROBPs. However, suitably extending the analysis of truly random restrictions by Chen et al. [CSV15] to the case of $\mathbf{AC}^0[\oplus]$ is nontrivial. We defer further discussion to Section 9.

[6] Again, this is only true up to sandwiching.

2 Preliminaries

2.1 Pseudorandomness primitives

Let U_n denote the uniform distribution over $\{0,1\}^n$. Suppose $\mathcal{C}$ is a class of functions $f\colon \{0,1\}^n \to \mathbb{R}$ and G is a distribution over $\{0,1\}^n$. We say that G *ε-fools* $\mathcal{C}$ if for every $f \in \mathcal{C}$,

$$|\,\mathbb{E}[f(G)] - \mathbb{E}[f(U_n)]| \le \varepsilon.$$

As two special cases, a *δ-biased* distribution is one that δ-fools parity functions, and a *γ-almost k-wise independent* distribution is one that γ-fools Boolean k-juntas [NN93, AGHP92]. An *ε-PRG* for $\mathcal{C}$ is a function $G\colon \{0,1\}^s \to \{0,1\}^n$ such that $G(U_s)$ ε-fools $\mathcal{C}$. As a shorthand, we will write $\mathbb{E}[f]$ to denote $\mathbb{E}[f(U_n)]$.

2.2 Read-once formulas

An $\mathbf{AC}^0$ *formula* on $\{0,1\}^n$ is a rooted tree in which each internal node ("gate") is labeled with $\wedge$ or $\vee$ and each leaf is labeled with a constant (0 or 1), a variable x_i, or its negation $\neg x_i$, where $i \in [n]$. Gates may have arbitrary fan-in. The formula computes a function $\phi\colon \{0,1\}^n \to \{0,1\}$ in the natural way. The *depth* of the formula is the length of the longest path from the output gate to a leaf. The formula is *read-once* if each variable x_i appears at most once. We make no assumptions about the order in which the variables appear. A *layered* $\mathbf{AC}^0$ formula is one in which the gates are arranged in alternating layers of $\wedge$ and $\vee$ gates. Any read-once $\mathbf{AC}^0$ formula can be simulated by a layered read-once $\mathbf{AC}^0$ formula of the same depth.

2.3 Random restrictions

A *restriction* is a string $x \in \{0,1,\star\}^n$. We define an associative *composition* operation on $\{0,1,\star\}^n$ by

$$(x \circ x')_i = \begin{cases} x_i & \text{if } x_i \neq \star \\ x'_i & \text{if } x_i = \star. \end{cases}$$

Conceptually, $x \circ x'$ corresponds to first restricting according to x and then further restricting according to x'. As a special case, if $x' \in \{0,1\}^n$, then $x \circ x' \in \{0,1\}^n$ is the string obtained by using x' to "fill in the $\star$ positions" of x. If $f\colon \{0,1\}^n \to \{0,1\}$ is a function and x is a restriction, we define the restricted function $(f|_x)\colon \{0,1\}^n \to \{0,1\}$ by

$$(f|_x)(x') = f(x \circ x').$$

We define R_n to be the distribution over $X \in \{0,1,\star\}^n$ in which the coordinates are independent, $\Pr[X_i = \star] = 1/2$, and $\Pr[X_i = 0] = \Pr[X_i = 1] = 1/4$. If H_1, H_2 are distributions over $\{0,1,\star\}^n$, we define $H_1 \circ H_2$ to be the distribution over $X \in \{0,1,\star\}^n$ obtained by drawing independent samples $X_1 \sim H_1, X_2 \sim H_2$ and composing them, $X = X_1 \circ X_2$. For a nonnegative integer s, we define

$$H^{\circ s} = \underbrace{H \circ H \circ \cdots \circ H}_{s \text{ times}}.$$

For example, $R_n^{\circ s}$ is a random restriction where each coordinate is $\star$ with probability 2^{-s} and the non-$\star$ positions are uniform random bits.

A restriction can be specified by two n-bit strings as follows. Define $\mathrm{Res}\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1,\star\}^n$ by

$$\mathrm{Res}(y,z)_i = \begin{cases} \star & \text{if } y_i = 1 \\ z_i & \text{if } y_i = 0. \end{cases}$$

In words, y indicates which positions have $\star$, and z specifies the bits in the non-$\star$ positions. Observe that $\mathrm{Res}(U_{2n}) \sim R_n$.

3 Our PRG Construction

The construction of our generator is by induction on the depth of the formula we wish to fool. For the base case of depth-2 formulas, we use the PRG by Gopalan et al. for read-once CNFs and DNFs [GMR+12]. For the inductive step, let $d \geq 2$ be arbitrary, let G_d be a random variable over $\{0,1\}^n$ that α-fools depth-d read-once $\mathbf{AC}^0$ formulas, and let $\varepsilon > 0$ be arbitrary. We will show how to ε-fool depth-$(d+1)$ formulas, assuming α is sufficiently small.

Step 1: XORing with small-bias and almost k-wise independence. Let G'_d be an independent copy of G_d. Sample T from a γ-almost k-wise independent distribution over $\{0,1\}^n$, and sample D from a δ-biased distribution over $\{0,1\}^n$, where the parameters γ, k, δ will be specified later. Define

$$\overline{G}_d = (G_d \oplus T, G'_d \oplus D) \in \{0,1\}^n \times \{0,1\}^n.$$

Step 2: Assigning most the inputs using $\overline{G}_d$. Define a pseudorandom restriction $H_d \in \{0,1,\star\}^n$ by

$$H_d = \mathrm{Res}(\overline{G}_d).$$

Since $\mathrm{Res}(U_{2n}) \sim R_n$, each coordinate of H_d is $\star$ with probability roughly $1/2$. For a parameter

$$s = O((d \log\log(n/\varepsilon)) \cdot \log\log n),$$

we will restrict according to $H_d^{\circ s}$, i.e., we will compose s independent copies of the restriction H_d.

Step 3: Assigning remaining inputs using the MRT PRG [MRT19]. We rely on a PRG by Meka, Reingold, and Tal for XORs of short ROBPs [MRT19]; we will discuss this in more detail in Section 7. Sample $G_{\mathrm{MRT}} \in \{0,1\}^n$ using this PRG. Our final PRG for depth-$(d+1)$ read-once $\mathbf{AC}^0$ is defined by

$$G_{d+1} = H_d^{\circ s} \circ G_{\mathrm{MRT}},$$

i.e., we use G_{MRT} to assign bits to all remaining $\star$-positions after restricting according to $H_d^{\circ s}$.

4 Pseudorandom Restrictions Preserve Expectation

Toward proving the correctness of our PRG, in this section, we will show that restricting a depth-$(d+1)$ formula using the distribution H_d approximately preserves the expectation of the formula.

The following lemma proved by Forbes and Kelley shows that bounded-width ROBPs behave nicely under pseudorandom restrictions that are defined by small biased distributions and almost k-wise independence. In the lemma, $\mathcal{L}(n,w;k)$ is defined to be the maximum of $\sum_{i=1}^{k} \sum_{S \subseteq [n], |S|=k} |\widehat{f}(S)|$ over all width-w ROBPs f, where $\widehat{f}(S)$ denotes the Fourier coefficient of f at S.

Lemma 4.1 (Lemma 7.2 from [FK18], rephrased)**.** *Let T and D be independent random variables over $\{0,1\}^n$, which are sampled respectively from a γ-almost k-wise independent distribution and a δ-biased distribution. Let $f\colon \{0,1\}^n \to \{0,1\}$ be a width-w arbitrarily-ordered ROBP. Then,*

$$\left| \mathop{\mathbb{E}}_{U\sim U_n}[f(U)] - \mathop{\mathbb{E}}_{\substack{T,D\\ V\sim U_n}}\left[f|_{\mathrm{Res}(T,D)}(V)\right] \right| \le \left(\sqrt{\delta}\cdot \mathcal{L}(n,w;k) + \left(\frac{1}{2}\right)^{k/2} + \sqrt{\gamma} \right)\cdot nw.$$

We are mainly interested in fooling $\mathbf{AC}^0$ formulas, but for the analysis, it will be helpful to consider NAND *formulas*, i.e., formulas in which each internal node is a NAND gate instead of an $\wedge$ gate or an $\vee$ gate. In Section 8, we will explain why it suffices to reason about NAND formulas.

Recall from Section 3 that $\overline{G}_d = (G_d \oplus T, G'_d \oplus D)$, where G_d and G'_d are independent random variables over $\{0,1\}^n$ that α-fool depth-d read-once formulas, T is sampled from a γ-almost k-wise independent distribution over $\{0,1\}^n$, and D is sampled from a δ-biased distribution over $\{0,1\}^n$. We will use the following simple application of the above lemma to our pseudorandom restriction $H_d = \mathrm{Res}(\overline{G}_d)$. Looking ahead, we will eventually choose $\varepsilon_0 = \varepsilon/\operatorname{poly}(\log\log(n/\varepsilon))$.

Lemma 4.2. *There exist constants $c_1, c_2, c_3 > 0$, such that for all positive integers n, d, for every $\varepsilon_0 > 0$, if we set*

$$k = c_1 \log(nd/\varepsilon_0), \quad \delta = \varepsilon_0 \cdot \left(\frac{c_2}{\log n}\right)^{-k(d+2)} \quad \text{and} \quad \gamma = \frac{c_3\varepsilon_0}{nd},$$

then H_d as defined above satisfies the following. For every depth-$(d+1)$ read-once NAND *formula $\phi\colon \{0,1\}^n \to \{0,1\}$,*

$$\left| \mathop{\mathbb{E}}_{U\sim U_n}[\phi(U)] - \mathop{\mathbb{E}}_{H_d, V\sim U_n}[\phi|_{H_d}(V)] \right| \le \varepsilon_0.$$

Proof. We start by noting that $G_d \oplus T$ and $G'_d \oplus D$ are independent, $G_d \oplus T$ is γ-almost k-wise independent, and $G'_d \oplus D$ is δ-biased. This is due to the fact that linear tests and k-juntas are closed under shifts.

The lemma is then an immediate corollary of Lemma 4.1, because every depth-$(d+1)$ read-once NAND formula can be computed by a width $d+2$ read-once branching program [CSV15], and $\mathcal{L}(n, d+2; k)$ is bounded by $O(\log n)^{k(d+2)}$ [CHRT18]. Thus

$$\left| \mathop{\mathbb{E}}_{U\sim U_n}[\phi(U)] - \mathop{\mathbb{E}}_{H_d, V\sim U_n}[\phi|_{H_d}(V)] \right| \le \left(\sqrt{\delta}\cdot O(\log n)^{k(d+2)} + \left(\frac{1}{2}\right)^{k/2} + \sqrt{\gamma} \right)\cdot n(d+2),$$

and it is easy to check that there are constants c_1, c_2, c_3 such that the right hand side is bounded by ε_0 for a choice of δ, γ, k as in the statement of the lemma. □

We get the following corollary about repeated applications of H_d immediately since depth-$(d+1)$ read-once formulas are closed under restrictions.

Corollary 4.3. *Let ϕ be a depth-$(d+1)$ read-once* NAND *formula over n variables. Let δ, k, γ be as in Lemma 4.2. Then, for every integer $t \ge 1$,*

$$\left| \mathop{\mathbb{E}}_{U\sim U_n}[\phi(U)] - \mathop{\mathbb{E}}_{H_d^{\circ t}, V\sim U_n}\left[\phi|_{H_d^{\circ t}}(V)\right] \right| \le \varepsilon_0 t.$$

5 Pseudorandom Restrictions Simplify Read-Once Formulas

In this section, we derandomize the analysis of Chen et al. [CSV15] and show that our pseudorandom restriction generator $H_d^{\circ t}$ simplifies depth-$(d+1)$ formulas, as we discussed in Section 1.2. We first introduce our progress measure.

Definition 5.1. *Given a read-once* NAND *formula ϕ, we let $\Delta(\phi)$ be the maximum fan-in of any gate in ϕ that is not the root.*

Our goal is to show that when X is sampled from $H_d^{\circ t}$ then a read-once formula ϕ is simplified in the sense that $\Delta(\phi|_X)$ is roughly $\sqrt{\Delta(\phi)}$, with high probability. We will show that $t = O(d \log\log(n/\varepsilon))$ is sufficient. Our analysis will closely follow the analysis by Chen et al. [CSV15] for truly random restrictions.

5.1 Truly random restrictions simplify depth-$(d-1)$ formulas

Chen, Steinke and Vadhan proved that biased read-once formulas collapse to a constant after a random restriction, with high probability [CSV15]. Looking ahead, we will eventually set $\theta = (\varepsilon/n)^{O(1)}$.

Lemma 5.2 ([CSV15], Lemma A.3)**.** *Let φ be a depth-d read-once* NAND *formula over n variables such that either $\mathbb{E}[\neg\varphi] \le \rho$ or $\mathbb{E}[\varphi] \le \rho$ for some $\rho \le \frac{1}{2}$. Then, for every $\theta \in (0, \frac{2}{n})$ and $p \le \frac{1}{(9\log(2\cdot 4^d n/\theta))^d}$ it holds that*

$$\Pr_{X \sim R_n^{\circ\lceil \log p^{-1}\rceil}}[\varphi|_X \text{ is not a constant}] \le 2p \cdot \rho \cdot (9\log(2\cdot 4^d n/\theta))^d + \theta.$$

We use Lemma 5.2 to prove the following variation; note that this lemma considers the case of several read-once formulas and analyzes the probability of collapsing to 1 instead of collapsing to any constant.

Lemma 5.3. *Let $\Phi = \{\phi_1, \ldots, \phi_k\}$ be a set of read-once* NAND *formulas over n variables, each of depth $d \le \log n$ and over disjoint subsets of n variables. Further, assume that for every $i \in [k]$, $\mathbb{E}[\neg\phi_i] \le \rho$ for some $\rho \le \frac{1}{2}$. Then, there exists a constant c such that for every $\theta \in (0, \frac{2}{n})$ and integer $t \ge cd\log\log(n/\theta)$,*

$$\Pr_{X\sim R_n^{\circ t}}[\forall \phi \in \Phi, \phi|_X \not\equiv 1] \le (2\rho + \theta)^k.$$

Proof. Consider some $\phi \in \Phi$ and let t be the smallest integer such that

$$2^{-t} \le \frac{1}{2(9\log(2\cdot 4^d n/\theta))^d},$$

and indeed $t = c\,(d\log\log(n/\theta) + d\log d)$ for some universal constant c. By Lemma 5.2,

$$\Pr_{X\sim R_n^{\circ t}}[\phi|_X \text{ is not a constant}] \le \rho + \theta.$$

Now,

$$\Pr_{X\sim R_n^{\circ t}}[\phi|_X \equiv 0] \le \mathbb{E}[\neg\phi] \le \rho,$$

so by the union bound

$$\Pr_{X\sim R_n^{\circ t}}[\phi|_X \not\equiv 1] \le 2\rho + \theta.$$

The lemma follows by the fact that each formula in Φ is over distinct variables and the coordinates of $R_n^{\circ t}$ are independent. $\square$

5.2 H_d simplifies depth-$(d-1)$ formulas

Ultimately, we are interested in the simplification of depth-$(d+1)$ formulas with respect to the $\Delta(\cdot)$ measure of progress. However, in this subsection, our goal is to prove that our iterated pseudorandom restriction $H_d^{\circ t}$ simplifies depth-$(d-1)$ formulas just as well as truly random restrictions up to an additive error. In this subsection, the notion of simplification is the event in the statement of Lemma 5.3.

Lemma 5.4. *Let $\Phi = \{\phi_1, \dots, \phi_k\}$ be a set of read-once* NAND *formulas over n variables, each of depth $d-1$ and over disjoint subsets of n variables. Then, for every integer $t \geq 1$,*

$$\Pr_{X \sim H_d \circ R_n^{\circ(t-1)}}[\forall \phi \in \Phi, \phi|_X \not\equiv 1] \leq \Pr_{X \sim R_n^{\circ t}}[\forall \phi \in \Phi, \phi|_X \not\equiv 1] + 2\alpha,$$

where α is the error of the PRG for depth-d read-once formulas underlying H_d.

Proof. Fix some restriction $v \in \{0,1,\star\}^n$. (Think of v as some fixing of $R_n^{\circ(t-1)}$.) Let $T_v \colon \{0,1\}^{2n} \to \{0,1\}$ be the predicate indicating that with respect to v, the given initial restriction does a poor job of simplifying Φ. That is,

$$T_v(y,z) = 1 \iff \forall \phi \in \Phi,\ \phi|_{\mathrm{Res}(y,z) \circ v} \not\equiv 1.$$

Claim 5.5. *For every $d \geq 2$, T_v can be computed by a depth-d read-once $\mathbf{AC}^0$ formula.*

Proof. We will prove, by induction on d, that for every $\phi \in \Phi$,

1. The test $\phi|_{\mathrm{Res}(y,z)\circ v} \not\equiv 1$ can be computed by a depth-d read-once $\mathbf{AC}^0$ formula with an $\wedge$ gate on top.
2. The test $\phi|_{\mathrm{Res}(y,z)\circ v} \not\equiv 0$ can be computed by a depth-d read-once $\mathbf{AC}^0$ formula with an $\vee$ gate on top.

The claim will then follow, as the "$\forall \phi \in \Phi$" part is simply an $\wedge$ over formulas with a top $\wedge$ gate and thus the two top layers can be collapsed to a single layer.

For $d = 2$, ϕ is of depth 1 and so is simply a NAND of variables or their negation, say of the literals $\ell_1, \dots, \ell_m$. Now,

$$\mathrm{NAND}(\ell_1, \dots, \ell_m) \not\equiv 1 \iff \bigwedge_{i \in [m]} (\ell_i \not\equiv 0),$$

and

$$\mathrm{NAND}(\ell_1, \dots, \ell_m) \not\equiv 0 \iff \bigvee_{i \in [m]} (\ell_i \not\equiv 1).$$

For each $b \in \{0,1\}$, let us express the condition $\ell_i \not\equiv b$ in terms of the inputs y and z to T_v.

- If ℓ_i is a variable x_i, then

$$x_i \not\equiv b \iff ((y_i = 1) \wedge (v_i \not\equiv b)) \vee ((y_i = 0) \wedge (z_i = \overline{b})).$$

  Now, v is fixed, so either $v_i \not\equiv b$ is the constant 0, in which case the formula amounts to $(y_i = 0) \wedge (z_i = \overline{b})$, or it is the constant 1, in which case the formula amounts to $(y_i = 1) \vee (z_i = \overline{b})$. Either way, this is a depth-1 read-once formula in terms of the inputs y and z to T_v.

- If ℓ_i is the negation $\neg x_i$ of some variable, then

$$\neg x_i \not\equiv b \iff ((y_i = 1) \wedge (v_i \not\equiv \overline{b})) \vee ((y_i = 0) \wedge (z_i = b))$$

  Again, by the same reasoning, the above is a depth-1 read-once formula, where the top gate is determined by the value of $v_i \not\equiv b$.

Thus, the predicate $\mathrm{NAND}(\ell_1, \ldots, \ell_m) \not\equiv 1$ can be tested by a depth-2 formula where the top gate is an $\wedge$, and the predicate $\mathrm{NAND}(\ell_1, \ldots, \ell_m) \not\equiv 0$ can be tested by a depth-2 formula where the top gate is an $\vee$.

Assume the claim holds for some $d \geq 2$ and let $\phi = \mathrm{NAND}(\varphi_1, \ldots, \varphi_m)$ be a read-once NAND formula of depth d, so each φ_i is a depth-$(d-1)$ read-once NAND formula. We already mentioned that

$$\mathrm{NAND}(\varphi_1, \ldots, \varphi_m) \not\equiv 1 \iff \bigwedge_{i \in [m]} (\varphi_i \not\equiv 0).$$

By the induction's hypothesis, the predicate $\varphi_i|_{\mathrm{Res}(y,z) \circ v} \not\equiv 0$ can be tested by a depth-d read-once $\mathbf{AC}^0$ formula with a top $\vee$ gate, so overall we get a depth-$(d+1)$ read-once $\mathbf{AC}^0$ formula with a top $\wedge$ gate. Similarly,

$$\mathrm{NAND}(\varphi_1, \ldots, \varphi_m) \not\equiv 0 \iff \bigvee_{i \in [m]} (\varphi_i \not\equiv 1).$$

Again, by our assumption, the predicate $\varphi_i|_{\mathrm{Res}(y,z) \circ v} \not\equiv 1$ can be tested by a depth-d read-once $\mathbf{AC}^0$ formula with a top $\wedge$ gate, so overall we get a depth-$(d+1)$ read-once $\mathbf{AC}^0$ formula with a top $\vee$ gate. $\square$

Recall from Section 3 the distribution

$$\overline{G}_d = (G_d \oplus T, G'_d \oplus D).$$

We shall later show:

Claim 5.6. *$\overline{G}_d$ (2α)-fools depth-d read-once $\mathbf{AC}^0$ formulas over $\{0,1\}^{2n}$.*

With the above claim in mind, and Claim 5.5, we are now ready to proceed with proving the lemma. We get that:

$$\Pr_{X \sim H_d}[\forall \phi \in \Phi, \phi|_{X \circ v} \not\equiv 1] = \Pr_{X \sim H_d}[T_v(X) = 1] \leq \Pr_{(Y,Z) \sim U_{2n}}[T_v(Y,Z) = 1] + 2\alpha.$$

A uniform (Y, Z) corresponds to a truly random restriction, so

$$\Pr_{X \sim H_d}[\forall \phi \in \Phi, \phi|_{X \circ v} \not\equiv 1] \leq \Pr_{X \sim R_n}[\forall \phi \in \Phi, \phi|_{X \circ v} \not\equiv 1] + 2\alpha.$$

As the above is true for every restriction v, obviously

$$\mathop{\mathbb{E}}_{V \sim R_n^{\circ(t-1)}}\left[\Pr_{X \sim H_d}[\forall \phi \in \Phi, \phi|_{X \circ V} \not\equiv 1]\right] \leq \mathop{\mathbb{E}}_{V \sim R_n^{\circ(t-1)}}\left[\Pr_{X \sim R_n}[\forall \phi \in \Phi, \phi|_{X \circ V} \not\equiv 1]\right] + 2\alpha,$$

so

$$\mathop{\mathbb{E}}_{X \sim H_d}\left[\Pr_{V \sim R_n^{\circ(t-1)}}[\forall \phi \in \Phi, \phi|_{X \circ V} \not\equiv 1]\right] \leq \Pr_{X \sim R_n^{\circ t}}[\forall \phi \in \Phi, \phi|_X \not\equiv 1] + 2\alpha,$$

which amounts to what we wanted to prove. All that is left is to prove Claim 5.6.

Proof of Claim 5.6. We start by noting that since depth-d read-once $\mathbf{AC}^0$ is closed under shifts, $G_d \oplus T$ and $G'_d \oplus D$ both α-fool depth-d read-once $\mathbf{AC}^0$.

We will next use the fact that depth-d read-once $\mathbf{AC}^0$ is closed under restrictions. Suppose $\phi\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is a depth-d read-once $\mathbf{AC}^0$ formula. We have

$$\left| \mathop{\mathbb{E}}_{U,V\sim U_n}[\phi(U,V)] - \mathop{\mathbb{E}}_{(X,Y)\sim \overline{G}_d}[\phi(X,Y)] \right| \leq$$

$$\left| \mathop{\mathbb{E}}_{V\sim U_n}\left[\mathop{\mathbb{E}}_{U\sim U_n}[\phi(U,V)] - \mathop{\mathbb{E}}_{X\sim G_d\oplus T}[\phi(X,V)] \right] \right| + \left| \mathop{\mathbb{E}}_{X\sim G_d\oplus T}\left[\mathop{\mathbb{E}}_{V\sim U_n}[\phi(X,V)] - \mathop{\mathbb{E}}_{Y\sim G'_d\oplus D}[\phi(X,Y)] \right] \right| \leq$$

$$\mathop{\mathbb{E}}_{V\sim U_n}\left| \mathop{\mathbb{E}}_{U\sim U_n}[\phi(U,V)] - \mathop{\mathbb{E}}_{X\sim G_d\oplus T}[\phi(X,V)] \right| + \mathop{\mathbb{E}}_{X\sim G_d\oplus T}\left| \mathop{\mathbb{E}}_{V\sim U_n}[\phi(X,V)] - \mathop{\mathbb{E}}_{Y\sim G'_d\oplus D}[\phi(X,Y)] \right| \leq 2\alpha,$$

where we used the fact that $G_d \oplus T$ and $G'_d \oplus D$ are independent and α-fool the formulas $\phi(\cdot, v)$ and $\phi(x, \cdot)$ respectively. □

□

Iterating H_d for t times, we get the following lemma. Roughly speaking, the proof is a hybrid argument of which Lemma 5.4 is a single step.

Lemma 5.7. *Let $\Phi = \{\phi_1, \ldots, \phi_k\}$ be a set of read-once* NAND *formulas over n variables, each of depth $d-1$ and over disjoint subsets of n variables. Then, for every integer $t \geq 1$,*

$$\Pr_{X\sim H_d^{\circ t}}[\forall \phi \in \Phi, \phi|_X \not\equiv 1] \leq \Pr_{X\sim R_n^{\circ t}}[\forall \phi \in \Phi, \phi|_X \not\equiv 1] + 2t\alpha,$$

where α is the error of the PRG for depth-d read-once $\mathbf{AC}^0$ formulas underlying H_d.

Proof. We prove the lemma by induction on t. The case of $t = 0$ is trivial. Now, assume that

$$\Pr_{X\sim H_d^{\circ(t-1)}}[\forall \phi \in \Phi, \phi|_X \not\equiv 1] \leq \Pr_{X\sim R_n^{\circ(t-1)}}[\forall \phi \in \Phi, \phi|_X \not\equiv 1] + 2(t-1)\alpha.$$

Thus,

$$\begin{aligned}
\Pr_{X\sim H_d^{\circ t}}[\forall \phi \in \Phi, \phi|_X \not\equiv 1] &= \mathop{\mathbb{E}}_{X_1\sim H_d}\left[\Pr_{X_2\sim H_d^{\circ(t-1)}}[\forall \phi \in \Phi, \phi|_{X_1\circ X_2} \not\equiv 1] \right] \\
&= \mathop{\mathbb{E}}_{X_1\sim H_d}\left[\Pr_{X_2\sim H_d^{\circ(t-1)}}[\forall \phi \in \Phi, (\phi|_{X_1})\,|_{X_2} \not\equiv 1] \right] \\
&\leq \mathop{\mathbb{E}}_{X_1\sim H_d}\left[\Pr_{X_2\sim R_n^{\circ(t-1)}}[\forall \phi \in \Phi, (\phi|_{X_1})\,|_{X_2} \not\equiv 1] \right] + 2(t-1)\alpha \\
&\leq \Pr_{X\sim R_n^{\circ t}}[\forall \phi \in \Phi, \phi|_X \not\equiv 1] + 2t\alpha.
\end{aligned}$$

The third transition used the induction's hypothesis and the last one is due to Lemma 5.4. □

Combining Lemma 5.7 with Lemma 5.3 we immediately get the following corollary.

Corollary 5.8. *Let $\Phi = \{\phi_1, \ldots, \phi_k\}$ be a set of* NAND *read-once formulas over n variables, each of depth $d-1$ and over disjoint subsets of n variables. Further, assume that $d \le \log n$ and that for every $i \in [k]$, $\mathbb{E}[\neg\phi_i] \le \rho$ for some $\rho \le \frac{1}{2}$. Then, there exists a constant c such that for every $\theta \in (0, \frac{2}{n})$ and integer $t \ge cd \log\log(n/\theta)$,*

$$\Pr_{X \sim H_d^{\circ t}} [\forall \phi \in \Phi, \phi|_X \not\equiv 1] \le (2\rho + \theta)^k + 2t\alpha,$$

where α is the error of the PRG for depth-d read-once $\mathbf{AC}^0$ formulas underlying H_d.

5.3 $H_d^{\circ t}$ simplifies depth-$(d+1)$ formulas

We are now ready to prove our main result for this section.

Lemma 5.9. *Let ϕ be a depth-$(d+1)$ read-once* NAND *formula over n variables where $d \le \log n$. Let $\varepsilon_0 > 0$ and let c be the constant guaranteed by Corollary 5.8. Further assume that $\theta \in (0, \frac{2}{n})$ is such that for every gate ψ in ϕ, possibly excluding the root, $\mathbb{E}[\neg\psi] \ge \theta$. Then, for every integer $t \ge cd \log\log(n/\theta)$ and every $\alpha \le \frac{\varepsilon_0^2}{8(dn)^2\sqrt{n}\log^2(1/\theta)t}$,*

$$\Pr_{X \sim H_d^{\circ t}} \left[\Delta(\phi|_X) \le 10\sqrt{\Delta(\phi)}\log^2(1/\theta)\right] \ge 1 - \varepsilon_0,$$

where the PRG for depth-d read-once $\mathbf{AC}^0$ formulas underlying H_d is instantiated with error α.

Note that we assume here that every gate in ϕ has a non-negligible probability of rejecting, which may not always be the case. Following Chen et al. [CSV15], in Section 6 we will get rid of that assumption by a sandwiching argument. The proof of Lemma 5.9 is based on an argument introduced by Gopalan et al. [GMR+12], later also used by Chen et al. [CSV15].

Proof. Let ψ be any gate in ϕ other than the root, so ψ is a depth-d read-once NAND formula. We shall partition its children Ψ according to their rejection probability. Namely, for every integer $0 \le i \le \log(1/\theta) - 1$ define

$$\Psi_i = \left\{\varphi \in \Psi : 2^i\theta \le \mathbb{E}[\neg\varphi] < 2^{i+1}\theta\right\}.$$

Note that if $\mathbb{E}[\neg\varphi] = 1$ then ψ is fixed to 1 so we can simply ignore it.

Let us fix some $0 \le i \le \log(1/\theta) - 1$ and consider the set of formulas Ψ_i. In hindsight, set the parameters

$$M = 5e\ln(1/\theta)\sqrt{\Delta(\phi)}$$

and

$$k = \left\lceil \frac{2}{\log\Delta(\phi)} \log\left(\frac{2dn\log(1/\theta)}{\varepsilon_0}\right)\right\rceil.$$

Write $\Psi_i = \{\varphi_1, \ldots, \varphi_w\}$. For every $j \in [w]$, let Y_j be the indicator for the event that φ_j is not identically 1 after a pseudorandom restriction, namely $\varphi_j|_X \not\equiv 1$. We wish to bound

$$\Pr\left[\sum_{j \in [w]} Y_j \ge M\right],$$

where the probability is taken over $X \sim H_d^{\text{ot}}$. Let

$$S_k(x_1, \ldots, x_w) = \sum_{I \subseteq [w], |I|=k} \prod_{i \in I} x_i$$

be the k-th elementary symmetric polynomial. Note that if $\sum_{j \in [w]} Y_j \geq M$ then $S_k(Y_1, \ldots, Y_w)$ is at least $\binom{M}{k}$, and so

$$\begin{aligned}
\Pr\left[\sum_{j \in [w]} Y_j \geq M\right] &\leq \frac{1}{\binom{M}{k}} \mathbb{E}[S_k(Y_1, \ldots, Y_w)] \\
&\leq \left(\frac{k}{M}\right)^k \sum_{I \subseteq [w], |I|=k} \Pr\left[\forall j \in I, Y_j = 1\right].
\end{aligned}$$

We know that $\mathbb{E}[\neg\varphi] \leq 2^{i+1}\theta$ and φ is a depth-$(d-1)$ NAND formula, so by Corollary 5.8 we get

$$\Pr\left[\sum_{j \in [w]} Y_j \geq M\right] \leq \left(\frac{k}{M}\right)^k \binom{w}{k} \left((2 \cdot 2^{i+1}\theta + \theta)^k + 2t\alpha\right). \tag{1}$$

Now,

Claim 5.10. *It holds that* $w \leq \frac{\ln(1/\theta)}{2^i\theta}$.

Proof. On the one hand,

$$\prod_{\varphi \in \Psi} \mathbb{E}[\varphi] = \mathbb{E}[\neg\psi] \geq \theta.$$

On the other hand,

$$\prod_{\varphi \in \Psi} \mathbb{E}[\varphi] \leq \prod_{\varphi \in \Psi_i} \mathbb{E}[\varphi] \leq (1 - 2^i\theta)^w \leq e^{-2^i w\theta}.$$

Combining the two gives the desired bound. □

Plugging in the above bound to Equation (1), we get

$$\begin{aligned}
\Pr\left[\sum_{j \in [w]} Y_j \geq M\right] &\leq \left(\frac{k}{M}\right)^k \left(\frac{we}{k}\right)^k \left((2 \cdot 2^{i+1}\theta + \theta)^k + 2t\alpha\right) \\
&\leq \left(\frac{ew \cdot (2^{i+2}\theta + \theta)}{M}\right)^k + 2\left(\frac{we}{M}\right)^k t\alpha \\
&\leq \left(\frac{5e\ln(1/\theta)}{M}\right)^k + 2\left(\frac{\Delta(\phi)e}{M}\right)^k t\alpha,
\end{aligned}$$

where for the second summand we only used the trivial fact that $w \leq \Delta(\phi)$.

Plugging in M, we achieve

$$\Pr\left[\sum_{j \in [w]} Y_j \geq M\right] \leq \frac{1}{\Delta(\phi)^{k/2}} + 2(\Delta(\phi))^{k/2} \cdot t\alpha. \tag{2}$$

As $k \geq \frac{2}{\log \Delta(\phi)} \log\left(\frac{2dn\log(1/\theta)}{\varepsilon_0}\right)$ we have that the first summand of Equation (2) is at most $\frac{\varepsilon_0}{2dn\log(1/\theta)}$. Also, the bound on α implies

$$\frac{2dn\log(1/\theta)}{\varepsilon_0} \leq \frac{\varepsilon_0}{4dn\log(1/\theta)t\alpha} \cdot \frac{1}{\sqrt{\Delta(\phi)}}$$

so

$$k \leq \frac{2}{\log \Delta(\phi)} \log\left(\frac{2dn\log(1/\theta)}{\varepsilon_0}\right) + 1 \leq \frac{2}{\log \Delta(\phi)} \log\left(\frac{\varepsilon_0}{4dn\log(1/\theta)t\alpha}\right)$$

and the second summand of Equation (2) is at most $\frac{\varepsilon_0}{2dn\log(1/\theta)}$ as well. Thus,

$$\Pr\left[\sum_{j\in[w]} Y_j \geq M\right] \leq \frac{\varepsilon_0}{dn\log(1/\theta)}.$$

Define $E_i = \sum_{j\in[w]} Y_j$. By union-bounding over $\Psi_0, \ldots, \Psi_{\log(1/\theta)-1}$ we get that

$$\Pr\left[\sum_{i=0}^{\log(1/\theta)-1} E_i \geq M\log(1/\theta)\right] \leq \sum_{i=0}^{\log(1/\theta)-1} \Pr\left[E_i\right] \leq \frac{\varepsilon_0}{dn}.$$

Another union bound over all possible ψ-s (at most dn of them) gives us the desired bound. $\square$

6 Ensuring Noticeable Chance of Rejecting

In Section 5, we showed that $H^{\circ t}$ simplifies formulas with high probability *under the assumption* that every gate rejects with noticeable probability. In this section, following Chen, Steinke, and Vadhan [CSV15], we will use a sandwiching argument to handle gates with negligible probability of rejecting. Our starting point is a helpful lemma implicit in the work of Chen et al. [CSV15]:

Lemma 6.1 ([CSV15])**.** *Suppose ϕ is a depth-d read-once* NAND *formula over n variables with $d \leq n$ and let $\varepsilon_0 > 0$. Define $\theta = \frac{\varepsilon_0^2}{4n^2}$. Then, there exist read-once* NAND *formulas ℓ_ϕ, u_ϕ with the following properties.*

1. *$\ell_\phi \leq \phi \leq u_\phi$ and $\mathbb{E}[u_\phi - \ell_\phi] \leq \varepsilon_0$.*
2. *The underlying tree structure of ℓ_ϕ is a subgraph of the underlying tree structure of ϕ, and the underlying tree structure of u_ϕ is a subgraph of the underlying tree structure of ϕ.*
3. *Every non-constant gate ψ in either ℓ_ϕ or u_ϕ satisfies $\mathbb{E}[\psi] \geq \theta$ and $\mathbb{E}[\neg\psi] \geq \theta$.*

Since Chen, Steinke, and Vadhan did not state Lemma 6.1 exactly as we have stated it here, for completeness, we include a proof of Lemma 6.1 in Appendix A.

The sandwiching formulas in Lemma 6.1 satisfy the hypothesis of Lemma 5.9, so after restricting according to $H^{\circ t}$, they simplify in the sense that Δ goes down by roughly a square root. We would like to apply $H^{\circ t}$ again to simplify the formulas even further. Unfortunately, after the first application of $H^{\circ t}$, the restricted formulas might no longer satisfy the hypothesis of Lemma 5.9. Therefore, before applying $H^{\circ t}$ the second time, we must apply Lemma 6.1 again. We will continue in this manner, alternately applying $H^{\circ t}$ to simplify and applying Lemma 6.1 to eliminate gates with negligible probability of rejecting. In this way, we will prove the following lemma.

Lemma 6.2. *Suppose ϕ is a depth-$(d+1)$ read-once* NAND *formula over n variables where $d \leq \log n$ and let $\varepsilon_0 > 0$. Assume the parameters $\alpha, k, \delta, \gamma$ underlying H_d satisfy the hypotheses of Lemma 5.9 and Lemma 4.2. Let θ be the value in Lemma 6.1, let t be as in Lemma 5.9, let $r = \lceil 3 \log\log n \rceil$, and let $s = rt$.*

Sample independent restrictions $X_1, \dots, X_r \sim H_d^{\circ t}$. For any such vector of restrictions $\vec{X}$, there exist depth-$(d+1)$ read-once NAND *formulas $\ell_{\phi,\vec{X}}, u_{\phi,\vec{X}}$ with the following properties.*

1. *(Bounding.) For every sample $\vec{X}$,*

$$\ell_{\phi,\vec{X}} \leq \phi|_{X_1 \circ \cdots \circ X_r} \leq u_{\phi,\vec{X}}.$$

2. *(Sandwiching.) For $U \sim U_n$ independent of $\vec{X}$,*

$$\mathop{\mathbb{E}}_{\vec{X},U} \left[u_{\phi,\vec{X}}(U) - \ell_{\phi,\vec{X}}(U) \right] \leq 3s\varepsilon_0.$$

3. *(Simplicity.) Let $\Delta_0 = 40^4 \log^8(2n/\varepsilon_0)$. Then,*

$$\Pr_{\vec{X}} \left[\Delta\left(\ell_{\phi,\vec{X}}\right) \leq \Delta_0 \text{ and } \Delta\left(u_{\phi,\vec{X}}\right) \leq \Delta_0 \right] \geq 1 - 2r\varepsilon_0.$$

Toward proving Lemma 6.2, fix a depth-$(d+1)$ read-once NAND formula ϕ, define $X_0 = \star^n$, and define $\ell^{(0)}_{\vec{X}} = u^{(0)}_{\vec{X}} = \phi$. Then, for $i < r$, inductively define

$$\ell^{(i+1)}_{\vec{X}} = \ell_{(\ell^{(i)}_{\vec{X}}|_{X_i})}.$$

That is, $\ell^{(i+1)}_{\vec{X}}$ is the lower sandwiching formula when Lemma 6.1 is applied to $\ell^{(i)}_{\vec{X}}|_{X_i}$. Similarly, define

$$u^{(i+1)}_{\vec{X}} = u_{(u^{(i)}_{\vec{X}}|_{X_i})},$$

i.e., $u^{(i+1)}_{\vec{X}}$ is the upper sandwiching formula when Lemma 6.1 is applied to $u^{(i)}_{\vec{X}}|_{X_i}$. Finally, define

$$\begin{aligned} \ell_{\phi,\vec{X}} &= \ell^{(r)}_{\vec{X}}|_{X_r} \\ u_{\phi,\vec{X}} &= u^{(r)}_{\vec{X}}|_{X_r}. \end{aligned}$$

Proof of Item 1 of Lemma 6.2. We show by induction on i that $\ell^{(i)}_{\vec{X}}|_{X_i} \leq \phi|_{X_1 \circ \cdots \circ X_i} \leq u^{(i)}_{\vec{X}}|_{X_i}$. In the base case $i = 0$, this is trivial. For the inductive step, we have

$$\begin{aligned} \ell^{(i+1)}_{\vec{X}}|_{X_{i+1}} &\leq \left(\ell^{(i)}_{\phi}|_{X_i}\right)|_{X_{i+1}} && \text{By Item 1 of Lemma 6.1} \\ &\leq (\phi|_{X_1 \circ \cdots \circ X_i})|_{X_{i+1}} && \text{By the induction's hypothesis} \\ &= \phi|_{X_1 \circ \cdots \circ X_{i+1}}. \end{aligned}$$

A completely analogous argument works for the upper bound as well. $\square$

Proof of Item 2 of Lemma 6.2. We show by induction on i that

$$\mathop{\mathbb{E}}_{\vec{X},U}\left[u_{\vec{X}}^{(i)}|_{X_i}(U) - \ell_{\vec{X}}^{(i)}|_{X_i}(U)\right] \leq (2t+2)i\varepsilon_0. \tag{3}$$

In the base case $i = 0$, the statement is trivial. For the inductive step, we have

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\vec{X},U}\left[u_{\vec{X}}^{(i+1)}|_{X_{i+1}}(U) - \ell_{\vec{X}}^{(i+1)}|_{X_{i+1}}(U)\right] \\
&\quad \leq \mathop{\mathbb{E}}_{\vec{X},U}\left[u_{\vec{X}}^{(i+1)}(U) - \ell_{\vec{X}}^{(i+1)}(U)\right] + 2t\varepsilon_0 && \text{By Corollary 4.3} \\
&\quad \leq \mathop{\mathbb{E}}_{\vec{X},U}\left[u_{\vec{x}}^{(i)}|_{X_i}(U) + \ell_{\vec{X}}^{(i)}|_{X_i}(U)\right] + (2t+2)\varepsilon_0 && \text{By Item 1 of Lemma 6.1} \\
&\quad \leq (2t+2)(i-1)\varepsilon_0 + (2t+2)\varepsilon_0. && \text{By the induction's hypothesis}
\end{aligned}$$

Finally, Item 2 of Lemma 6.2 follows from Equation (3) by plugging-in $i = r$ and as $s = rt$. □

Proof of Item 3 of Lemma 6.2. By construction, for every $i \geq 1$, the formula $\ell_{\vec{X}}^{(i)}$ and the formula $u_{\vec{X}}^{(i)}$ both have the property that every gate ψ satisfies $\mathbb{E}[\neg\psi] \geq \theta$, where

$$\theta = \frac{\varepsilon_0^2}{4n^2}.$$

Furthermore, as the restrictions are independent, X_i is independent of $\left(\ell_{\vec{X}}^{(i)}, u_{\vec{X}}^{(i)}\right)$. Therefore, by Lemma 5.9,

$$\Pr_{\vec{X}}\left[\Delta\left(\ell_{\vec{X}}^{(i)}|_{X_i}\right) > 10\sqrt{\Delta\left(\ell_{\vec{X}}^{(i)}\right)} \cdot \log^2(1/\theta)\right] \leq \varepsilon_0,$$

and

$$\Pr_{\vec{X}}\left[\Delta\left(u_{\vec{X}}^{(i)}|_{X_i}\right) > 10\sqrt{\Delta\left(u_{\vec{X}}^{(i)}\right)} \cdot \log^2(1/\theta)\right] \leq \varepsilon_0.$$

By the union bound, we may assume that none of these bad events occur and accumulate an error of $2\varepsilon_0$ for every restriction. Based on this assumption, we now show by induction on i that

$$\Delta\left(\ell_{\vec{X}}^{(i)}|_{X_i}\right) \leq \max\left\{10^4 \log^8(1/\theta), n^{(3/4)^i}\right\}, \tag{4}$$

and

$$\Delta\left(u_{\vec{X}}^{(i)}|_{X_i}\right) \leq \max\left\{10^4 \log^8(1/\theta), n^{(3/4)^i}\right\}. \tag{5}$$

The base case $i = 0$ follows from the trivial bound $\Delta(\phi) \leq n$. Now the inductive step. We have

$$\begin{aligned}
\Delta\left(\ell_{\vec{X}}^{(i+1)}|_{X_{i+1}}\right) &\leq 10\sqrt{\Delta\left(\ell_{\vec{X}}^{(i+1)}\right)} \cdot \log^2(1/\theta) && \text{By our assumption} \\
&\leq 10\sqrt{\Delta\left(\ell_{\vec{X}}^{(i)}|_{x_i}\right)} \cdot \log^2(1/\theta) && \text{By Item 2 of Lemma 6.1} \\
&\leq 10\sqrt{\max\left\{10^4 \log^8(1/\theta), n^{(3/4)^i}\right\}} \cdot \log^2(1/\theta) && \text{By the induction's hypothesis}
\end{aligned}$$

Now we have two cases. First, suppose $n^{(3/4)^i} \le 10^4 \log^8(1/\theta)$. Then the bound becomes

$$\begin{aligned}\Delta\left(\ell_{\vec{X}}^{(i+1)}\right) &\le 10\sqrt{10^4 \log^8(1/\theta)} \cdot \log^2(1/\theta) \\ &= 10^3 \log^6(1/\theta) \\ &\le 10^4 \log^8(1/\theta),\end{aligned}$$

completing the proof of Equation (4) in this case. Now, suppose instead that $10^4 \log^8(1/\theta) < n^{(3/4)^i}$. Then the bound becomes

$$\begin{aligned}\Delta\left(\ell_{\vec{X}}^{(i+1)}\right) &\le 10\sqrt{n^{(3/4)^i}} \cdot \log^2(1/\theta) \\ &\le \sqrt{n^{(3/4)^i}} \cdot (n^{(3/4)^i})^{1/4} \\ &= n^{(3/4)^{i+1}},\end{aligned}$$

once again completing the proof of Equation (4). The proof of Equation (5) is completely analogous and we omit it. Item 3 of Lemma 6.2 follows because by our choice of r, $n^{(3/4)^r} \le 2$, and by the definition of θ,

$$10^4 \log^8(1/\theta) = 40^4 \log^8(2n/\varepsilon_0).$$

□

7 Fooling Formulas when Δ is Small

Recall from Section 3 that our pseudorandom distribution for depth-$(d+1)$ read-once formulas is

$$H_d^{\circ s} \circ G_{\mathrm{MRT}}.$$

So far, we have shown that up to sandwiching, applying $H_d^{\circ s}$ substantially simplifies the formula with high probability while approximately preserving its expectation (Lemma 6.2). It remains to show that G_{MRT} fools these simpler formulas. Meka, Reingold, and Tal studied the problem of fooling XORs of short ROBPs and achieved the following parameters.

Theorem 7.1 ([MRT19]). *For any positive integers n, w, b and any $\varepsilon_0 > 0$ there is an explicit PRG that ε_0-fools all functions $f\colon \{0,1\}^n \to \{\pm 1\}$ of the form*

$$f(x) = \prod_{i=1}^{m} g_i(x),$$

where $g_1, \ldots, g_m : \{0,1\}^n \to \{\pm 1\}$ are defined over disjoint variable sets of size at most b and each g_i can be computed by an arbitrarily ordered width-w ROBP. The seed length of the PRG is

$$\log(n/\varepsilon_0) \cdot O(\log b + \log\log(n/\varepsilon_0))^{2w+2}.$$

It immediately follows that we can fool formulas when Δ is small with the following parameters.

Corollary 7.2. *For any integers n, d, Δ_0 and any $\varepsilon_0 > 0$, there is an explicit distribution G_{MRT} that ε_0-fools depth-d read-once* NAND *formulas ϕ satisfying $\Delta(\phi) \le \Delta_0$ that can be sampled using*

$$\log(n/\varepsilon_0) \cdot O(d \log \Delta_0 + \log\log(n/\varepsilon_0))^{2d+2}$$

truly random bits.

Proof. Write $\phi = \text{NAND}(\varphi_1, \ldots, \varphi_m)$. Then $\neg\phi = \wedge_{i=1}^m \varphi_i$. Applying the Fourier expansion of the m-input $\wedge$ function gives

$$\neg\phi = \sum_{S \subseteq [m]} \frac{(-1)^{|S|}}{2^m} \cdot \prod_{i \in S} (-1)^{\varphi_i}.$$

Since $\sum_S \left| \frac{(-1)^{|S|}}{2^m} \right| = 1$, it suffices to fool each function $\prod_{i \in S} (-1)^{\varphi_i}$ separately.

Since $\Delta(\phi) \leq \Delta_0$, each φ_i depends on at most Δ_0^{d-1} variables. Since ϕ is read-once, the φ_i-s depend on disjoint sets of variables. Since each φ_i is a depth-$(d-1)$ read-once NAND formula, it can be computed by a width-d ROBP under some ordering of the variables [CSV15]. Applying Theorem 7.1 completes the proof, since fooling ϕ is equivalent to fooling $\neg\phi$. □

8 Putting Everything Together: Proof of Theorem 1.1

To prove the correctness of our PRG, we first need to justify the fact that our analysis has so far focused on NAND formulas whereas our main result governs $\mathbf{AC}^0$ formulas, i.e., formulas over the $\{\wedge, \vee, \neg\}$ basis.

Lemma 8.1. *For any layered read-once $\mathbf{AC}^0$ formula ϕ, either ϕ or $\neg\phi$ can be computed by a read-once* NAND *formula with the same underlying tree structure as ϕ.*

Proof. We proceed by induction on the depth d of ϕ to show that if the output gate of ϕ is $\vee$, then ϕ can be computed by a read-once NAND formula with the same underlying tree structure as ϕ. In the base case $d = 1$, we have $\phi = \vee_{i=1}^m \ell_i$, where each ℓ_i is a literal. Then we can also write

$$\phi = \text{NAND}(\neg\ell_1, \ldots, \neg\ell_m).$$

Now, for the inductive step, assume $\phi = \vee_{i=1}^m \varphi_i$, where each φ_i is a depth-d read-once formula with output gate $\wedge$. Then once again,

$$\phi = \text{NAND}(\neg\varphi_1, \ldots, \neg\varphi_m).$$

By moving $\neg$ gates downward, $\neg\varphi_i$ can be converted to a depth-d read-once formula with output gate $\vee$ without altering its underlying tree structure. Applying the induction's hypothesis completes the proof. Finally, the lemma follows, because if the output gate of ϕ is $\wedge$, then $\neg\phi$ can be computed by a read-once formula with the same underlying tree structure with output gate $\vee$. □

Conversely, any read-once NAND formula can be straightforwardly simulated by a layered read-once $\mathbf{AC}^0$ formula with the same underlying tree structure. We are now ready to complete the analysis of our PRG.

Proof of Theorem 1.1. Recall that our PRG is $G_{d+1} = H_d^{\circ s} \circ G_{\text{MRT}}$.

Parameters. Assume $d \leq \log\log(n/\varepsilon)$. (Otherwise, Theorem 1.1 follows already from the work of Forbes and Kelley [FK18].) Let c be the constant from Lemma 5.3. Let $r = \lceil 3 \log\log n \rceil$, and define

$$\varepsilon_0 = \frac{\varepsilon}{10r \cdot cd \log\log(n/\varepsilon)}.$$

Let $\theta = \frac{\varepsilon_0^2}{4n^2}$. Let $t = cd\lceil \log\log(n/\theta) \rceil$ (without loss of generality, take c to be an integer), and let $s = tr$. Let $\alpha = \varepsilon^4/n^3$; this is small enough to satisfy the hypothesis of Lemma 5.9. Let k, δ, γ be the values required by Lemma 4.2.

Correctness. Let ϕ be a depth-$(d+1)$ read-once $\mathbf{AC}^0$ formula. We can straightforwardly make ϕ a *layered* read-once $\mathbf{AC}^0$ formula without changing its depth. Since fooling ϕ is equivalent to fooling $\neg\phi$, by Lemma 8.1, we may assume that ϕ is a depth-$(d+1)$ read-once NAND formula. Since $s = tr$, we can write $H_d^{\circ s} = (H_d^{\circ t})^{\circ r}$. Consider drawing independent samples $X_1, \ldots, X_r \sim H_d^{\circ t}$. Let $\ell_{\phi,\vec{X}}, u_{\phi,\vec{X}}$ be the formulas guaranteed to us by Lemma 6.2. For brevity, let $G = G_{\mathrm{MRT}}$, and let $U \sim U_n$ be independent of G and $H_d^{\circ s}$. Let E be the high-probability event of Item 3 of Lemma 6.2, so whether E occurs depends only on $\vec{X}$. Then,

$$\begin{aligned}
\mathop{\mathbb{E}}_{G_{d+1}}[\phi(G_{d+1})] &= \mathop{\mathbb{E}}_{\vec{X}}\left[\mathop{\mathbb{E}}_{G}[\phi|_{X_1\circ\cdots\circ X_r}(G)]\right] && \text{By the definition of } G_{d+1} \\
&\le \mathop{\mathbb{E}}_{\vec{X}}\left[\mathop{\mathbb{E}}_{G}[u_{\phi,\vec{X}}(G)]\right] && \text{By Item 1 of Lemma 6.2} \\
&\le \mathop{\mathbb{E}}_{\vec{X}}\left[\mathop{\mathbb{E}}_{G}[u_{\phi,\vec{X}}(G)] \;\middle|\; E\right] + \Pr_{\vec{X}}[\neg E] \\
&\le \mathop{\mathbb{E}}_{\vec{X}}\left[\mathop{\mathbb{E}}_{U}[u_{\phi,\vec{X}}(U)] + \varepsilon_0 \;\middle|\; E\right] + \Pr_{\vec{X}}[\neg E] && \text{By Corollary 7.2} \\
&\le \mathop{\mathbb{E}}_{\vec{X}}\left[\mathop{\mathbb{E}}_{U}[u_{\phi,\vec{X}}(U)] + \varepsilon_0\right] + 2\Pr_{\vec{X}}[\neg E] \\
&\le \mathop{\mathbb{E}}_{\vec{X},U}[u_{\phi,\vec{X}}(U)] + (1+2r)\varepsilon_0 && \text{By Item 3 of Lemma 6.2} \\
&\le \mathop{\mathbb{E}}_{\vec{X},U}[\phi|_{X_1\circ\cdots\circ X_r}(U)] + (1+2r+3s)\varepsilon_0 && \text{By Item 2 of Lemma 6.2} \\
&\le \mathbb{E}[\phi] + (1+2r+4s)\varepsilon_0 && \text{By Corollary 4.3.}
\end{aligned}$$

A completely analogous argument handles the lower bound. To complete the proof of correctness, we verify that with our choice of parameters, the error is bounded by ε:

$$(1+2r+4s)\varepsilon_0 \le 5s\varepsilon_0 \le \frac{1+\log\log(n/\theta)}{2\log\log(n/\varepsilon)} \cdot \varepsilon \le \varepsilon.$$

Seed length. Let $q(n,d,\varepsilon)$ denote the seed length of our ε-PRG for depth-d read-once $\mathbf{AC}^0$. We will prove by induction on d that

$$q(n,d,\varepsilon) \le \log(n/\varepsilon) \cdot (Cd\log\log(n/\varepsilon))^{2d+2}, \tag{6}$$

where C is an absolute constant to be specified later.

In the base case $d = 2$, our PRG is just the PRG by Gopalan et al. [GMR+12], which has seed length $C_1 \log(n/\varepsilon)(\log\log(n/\varepsilon))^3$ for some absolute constant C_1. Since $2d+2 > 3$, we can ensure that Equation (6) holds by choosing $C > C_1$.

Now, for the inductive step, fix $d \ge 2$ and consider G_{d+1}. We can divide the seed length of G_{d+1} into three components.

- (The inductive seed length.) To sample from $H_d^{\circ s}$, we must draw $2s$ independent samples from G_d. The number of truly random bits required for this process is bounded by $2s \cdot q(n,d,\alpha)$. There is an absolute constant C_2 so that $s \le (C_2 d\log\log(n/\varepsilon))^2$. By induction and our choice of $\alpha = \varepsilon^4/n^3$, the number of truly random bits for this component, q_1, is bounded by

$$q_1 \le 8\log(n/\varepsilon) \cdot (Cd)^{2d+2} \cdot (2+\log\log(n/\varepsilon))^{2d+2} \cdot s.$$

To handle the additive 2 term in the middle, we can bound

$$\begin{aligned}(2+\log\log(n/\varepsilon))^{2d+2} &= (\log\log(n/\varepsilon))^{2d+2}\cdot\left(1+\frac{2}{\log\log(n/\varepsilon)}\right)^{2d+2}\\ &\leq (\log\log(n/\varepsilon))^{2d+2}\cdot\exp\left(\frac{4d+4}{\log\log(n/\varepsilon)}\right)\\ &\leq e^8,\end{aligned}$$

since we assumed $d \leq \log\log(n/\varepsilon)$. Therefore,

$$\begin{aligned}q_1 &\leq 8\cdot e^8\cdot\log(n/\varepsilon)\cdot(Cd\log\log(n/\varepsilon))^{2d+2}\cdot(C_2 d\log\log(n/\varepsilon))^2\\ &\leq \frac{1}{3}\log(n/\varepsilon)\cdot(C(d+1)\log\log(n/\varepsilon))^{2(d+1)+2}\end{aligned}$$

as long as we choose $C > C_2$.

- (The seed length for D and T.) To sample from $H_d^{\circ s}$, we must also draw $2s$ independent samples from D and T. Using standard constructions [NN93, AGHP92], the number of truly random bits required for this process, q_2, is $2s\cdot O(k+\log(n/\delta)+\log(1/\gamma))$. For some absolute constant C_3, by our choices of k,δ,γ, this is bounded by

$$\begin{aligned}q_2 &\leq C_3 d^2\log(n/\varepsilon)\log\log(n/\varepsilon)\log\log n\\ &\leq \frac{1}{3}\log(n/\varepsilon)\cdot(C(d+1)\log\log(n/\varepsilon))^{2(d+1)+2},\end{aligned}$$

provided $C > C_3$.

- (The seed length for the MRT generator.) Because of our choices for the parameters ε_0 and Δ_0, there is an absolute constant C_4 such that in the construction of G_{d+1}, the seed length q_3 of the distribution G_{MRT} from Corollary 7.2 satisfies

$$q_3 \leq \log(n/\varepsilon)\cdot(C_4(d+1)\log\log(n/\varepsilon))^{2(d+1)+2}.$$

Choosing $C > C_4$ ensures

$$q_3 \leq \frac{1}{3}\log(n/\varepsilon)\cdot(C(d+1)\log\log(n/\varepsilon))^{2(d+1)+2}.$$

Summing up q_1, q_2, q_3 completes the proof of Equation (6).

Explicitness. Our PRG construction combines explicit PRGs in a straightforward way, so it is explicit as well. □

9 Fooling Read-Once $\mathbf{AC}^0[\oplus]$ Formulas With a Few Parity Gates

In this section, as outlined in Section 1.3, we prove Theorem 1.2, which extends our main theorem to the case of $\mathbf{AC}^0[\oplus]$ formulas with a bounded number of parity gates. (An $\mathbf{AC}^0[\oplus]$ *formula* is defined just like an $\mathbf{AC}^0$ formula except that the gates may be labled $\wedge$, $\vee$, or $\oplus$.) The main challenge in proving Theorem 1.2 is that the sandwiching argument from Section 6 does not easily generalize. The trouble is that the parity function is not monotone, so it does not compose well with sandwiching formulas. This difficulty already arises in the special case of $\mathrm{PARITY} \circ \mathbf{AC}^0$, i.e., the case that the root gate is a parity gate and there are no other parity gates. Instead of true sandwiching formulas, we merely get the following: For every read-once $\mathrm{PARITY} \circ \mathbf{AC}^0$ formula ϕ, there is a $\mathrm{PARITY} \circ \mathbf{AC}^0$ formula $\widetilde{\phi}$ in which every gate rejects with non-negligible probability; this formula $\widetilde{\phi}$ *approximates* ϕ in the sense that

$$\Pr_{X \sim U_n}\left[\phi(X) = \widetilde{\phi}(X)\right] \approx 1.$$

This does not straightforwardly imply correctness of our PRG, because it says nothing about the expectation of ϕ under our *pseudorandom* distribution.

Briefly, to resolve this difficulty, we also design an auxiliary $\mathbf{AC}^0$ formula T_ϕ that certifies that most points x satisfy $\phi(x) = \widetilde{\phi}(x)$. Since T_ϕ is itself fooled by our PRG, $\widetilde{\phi}$ must be a good approximation of ϕ under our pseudorandom distribution as well as the uniform distribution, i.e.,

$$\Pr_{X \sim G_d}\left[\phi(X) = \widetilde{\phi}(X)\right] \approx 1.$$

This condition is a suitable alternative to the sandwiching condition. (A similar approach has been taken in several other works, e.g., [Bra09, CZ16, MRT19].)

9.1 Special case: Read-once $\mathrm{PARITY} \circ \mathbf{AC}^0$

Toward proving Theorem 1.2, we begin by considering read-once formulas of the form $\mathrm{PARITY} \circ \mathbf{AC}^0$. Fix any positive integers n, d and any $\varepsilon_1 > 0$. Let $H_d^{\circ s} \circ G_{\mathrm{MRT}}$ be our ε_1-PRG for depth-$(d+1)$ read-once $\mathbf{AC}^0$ formulas used to prove Theorem 1.1, but with different values for the parameters k, δ, γ (we will explain the changes later). We will prove the following.

Lemma 9.1 (Fooling read-once $\mathrm{PARITY} \circ \mathbf{AC}^0$)**.** *Let $\phi = \bigoplus_{j=1}^m \phi_j$, where each ϕ_j is a depth-d read-once $\mathbf{AC}^0$ formula, $\phi_1, \ldots, \phi_m$ are on disjoint variable sets, and ϕ is defined over $\{0,1\}^n$. Then $H_d^{\circ 2s} \circ G_{\mathrm{MRT}}$ fools ϕ with error $n^2 \varepsilon_1$.*

Note that the PRG in Lemma 9.1 applies twice as many independent copies of H_d as the PRG in the proof of Theorem 1.1. Note also that the PRG G_d that underlies H_d is merely assumed to fool depth-d read-once $\mathbf{AC}^0$ formulas (i.e., without any parity gates).

In the remainder of this subsection, we sketch the proof of Lemma 9.1 by reviewing the proof of Theorem 1.1 and making the necessary alterations.

9.1.1 H_d still preserves the expectation

The analogue of Corollary 4.3 still holds in the $\mathrm{PARITY} \circ \mathbf{AC}^0$ setting, with suitable changes to the constants:

Lemma 9.2. *There exist absolute constants $c_1', c_2', c_3' > 0$, such that if we set*

$$k = c_1' \log(nd/\varepsilon_0), \qquad \delta = \varepsilon_0 \cdot \left(\frac{c_2'}{\log n}\right)^{-k(2d+2)}, \qquad \text{and} \qquad \gamma = \frac{c_3' \varepsilon_0}{nd},$$

then H_d satisfies the following. Let $\phi = \bigoplus_{j=1}^m \phi_j$, where each ϕ_j is a depth-d read-once NAND *formula, $\phi_1, \ldots, \phi_m$ are on disjoint variable sets, and ϕ is defined on $\{0,1\}^n$. Then, for every integer $t \geq 1$,*

$$\left|\mathop{\mathbb{E}}_{U \sim U_n}[\phi(U)] - \mathop{\mathbb{E}}_{H_d^{\circ t}, V \sim U_n}\left[\phi|_{H_d^{\circ t}}(V)\right]\right| \leq \varepsilon_0 t.$$

Proof sketch. The argument is essentially the same as the proof of Corollary 4.3. The only change is the width bound. The parity function can be computed by a width-2 ROBP and each ϕ_j can be simulated by a width-$(d+1)$ ROBP, so we can simulate ϕ by a width-$(2d+2)$ ROBP. $\square$

9.1.2 $H_d^{\circ t}$ still simplifies formulas where each gate rejects with noticeable probability

Once again, for a formula ϕ as in Lemma 9.2, we define $\Delta(\phi)$ to be the maximum fan-in of any gate other than the root. The analogue of Lemma 5.9 also still holds in this setting:

Lemma 9.3. *Let ϕ be as in Lemma 9.2. Assume $d \leq \log n$, let $\varepsilon_0 > 0$, and let c be the constant guaranteed by Corollary 5.8. Further assume that $\theta \in (0, \frac{2}{n})$ is such that for every gate ψ in ϕ, possibly excluding the root, $\mathbb{E}[\neg\psi] \geq \theta$. Then, for every integer $t \geq cd \log\log(n/\theta)$ and every $\alpha \leq \frac{\varepsilon_0^2}{8(dn)^2\sqrt{n}\log^2(1/\theta)t}$,*

$$\Pr_{X \sim H_d^{\circ t}}\left[\Delta(\phi|_X) \leq 10\sqrt{\Delta(\phi)}\log^2(1/\theta)\right] \geq 1 - \varepsilon_0,$$

where the PRG for depth-d read-once formulas underlying H_d is instantiated with error α.

Proof. The proof of Lemma 5.4 still works in this setting, because if ψ is a gate other than the root, then the subformula rooted at ψ is a read-once NAND formula of depth at most d. $\square$

9.1.3 Ensuring noticeable chance of rejecting

As discussed at the beginning of this section, we are not able to generalize Lemma 6.1 to the PARITY $\circ$ $\mathbf{AC}^0$ setting. However, in the original setting of NAND formulas, we can strengthen Lemma 6.1 by obtaining a read-once $\mathbf{AC}^0$ formula that certifies that the sandwiching formulas are good approximations. Here, for simplicity and because it is sufficient, we focus on the lower sandwiching formula:

Lemma 9.4. *Let ϕ, ε_0, and ℓ_ϕ be as in Lemma 6.1. There is a depth-d read-once $\mathbf{AC}^0$ formula $T_\phi^\ell\colon \{0,1\}^n \to \{0,1\}$ such that $\mathbb{E}[T_\phi^\ell] \geq 1 - \varepsilon_0$, and for every $x \in \{0,1\}^n$, if $T_\phi^\ell(x) = 1$, then*

$$\ell_\phi(x) = \phi(x).$$

We defer the proof of Lemma 9.4 to Appendix A, where we prove the generalization involving both the lower and the upper sandwiching formulas (Lemma A.1). Just like in Section 6, we must alternately apply Lemma 9.4 to ensure non-negligible chance of rejection and Lemma 9.3 to argue that the formula simplifies. The following lemma is analogous to Lemma 6.2.

Lemma 9.5. *Let ϕ be as in Lemma 9.2. Assume the parameters $\alpha, k, \delta, \gamma$ underlying H_d satisfy the hypotheses of Lemma 9.3 and Lemma 9.2. Let θ be the value in Lemma 6.1, let t be as in Lemma 9.3, let $r = \lceil 3 \log\log n \rceil$, and let $s = rt$.*

Sample independent restrictions $X_1, \dots, X_r \sim H_d^{\circ t}$. For any such vector of restrictions $\vec{X}$, there is a formula $\widetilde{\phi}_{\vec{X}} = \bigoplus_{j=1}^m \widetilde{\phi}_j$, where each $\widetilde{\phi}_j$ is a depth-d read-once NAND formula and $\widetilde{\phi}_1, \dots, \widetilde{\phi}_m$ are on disjoint variable sets, and there is a function $T_{\phi,\vec{X}}\colon \{0,1\}^n \to \{0,1\}$ with the following properties.

1. *(Success indication.) For every sample $\vec{X}$ and every point $x \in \{0,1\}^n$, if $T_{\phi,\vec{X}}(x) = 1$, then*

$$\widetilde{\phi}_{\vec{X}}(x) = (\phi|_{X_1 \circ \cdots \circ X_r})(x).$$

2. *(Approximation.) If G ε_1-fools depth-d read-once $\mathbf{AC}^0$ formulas and is independent of $\vec{X}$, then*
$$\mathop{\mathbb{E}}_{\vec{X},G}\left[T_{\phi,\vec{X}}(G)\right] \geq 1 - mr(\varepsilon_1 + (s+1)\varepsilon_0).$$

3. *(Simplicity.) Let $\Delta_0 = 40^4 \log^8(2n/\varepsilon_0)$. Then,*
$$\Pr_{\vec{X}}\left[\Delta\left(\widetilde{\phi}_{\vec{X}}\right) \leq \Delta_0\right] \geq 1 - r\varepsilon_0.$$

The proof of Lemma 9.5 is similar to the proof of Lemma 6.2, and we defer it to Appendix B.

9.1.4 $G_{\mathbf{MRT}}$ still fools formulas when Δ is small

The analogue of Corollary 7.2 still holds in the PARITY $\circ$ $\mathbf{AC}^0$ setting:

Lemma 9.6. *Fix any positive integers n, d, Δ_0 and any $\varepsilon_0 > 0$. Let ϕ be as in Lemma 9.2, assume $\Delta(\phi) \leq \Delta_0$, and let G_{MRT} be as in Corollary 7.2. Then G_{MRT} fools ϕ with error $\varepsilon_0/2$.*

Proof sketch. We can write
$$\phi = \bigoplus_{j=1}^{m} \phi_j = \frac{1}{2} - \frac{1}{2}\prod_{j=1}^{m}(-1)^{\phi_j}.$$
The rest of the argument is the same as in the proof of Corollary 7.2. □

9.1.5 Putting everything together for PARITY $\circ$ $\mathbf{AC}^0$

Proof sketch of Lemma 9.1. We can straightforwardly make each ϕ_j a layered read-once formula without changing its depth. By Lemma 8.1, either ϕ_j or $\neg\phi_j$ can be computed by a read-once NAND formula with the same underlying tree structure. Furthermore, $\neg$ gates can be pushed upward through $\oplus$ gates. Therefore, since fooling ϕ is the same as fooling $\neg\phi$, we may simply assume that $\phi_1, \ldots, \phi_m$ are NAND formulas.

Since $s = tr$, we can write $H_d^{\circ 2s} = (H_d^{\circ t})^{\circ r} \circ H_d^{\circ s}$. Consider drawing independent samples $X_1, \ldots, X_r \sim H_d^{\circ t}, Y \sim H_d^{\circ s}$. Let $\widetilde{\phi}_{\vec{X}}, T_{\phi,\vec{X}}$ be the functions guaranteed to us by Lemma 9.5. For brevity, let $G = G_{\mathrm{MRT}}$, and let $U \sim U_n$, all independent of $X_1, \ldots, X_r, Y$. Let E be the high-probability event of Item 3 of Lemma 9.5, so whether E occurs depends only on $\vec{X}$. Then,
$$\mathop{\mathbb{E}}_{H_d^{\circ 2s},G}[\phi(H_d^{\circ 2s} \circ G)] = \mathop{\mathbb{E}}_{\vec{X},Y,G}\left[\phi|_{X_1 \circ \cdots \circ X_r}(Y \circ G)\right].$$

By Item 2 of Lemma 9.5,
$$\left|\mathop{\mathbb{E}}_{\vec{X},Y,G}\left[\phi|_{X_1 \circ \cdots \circ X_r}(Y \circ G)\right] - \mathop{\mathbb{E}}_{\vec{X},Y,G}\left[\widetilde{\phi}_{\vec{X}}(Y \circ G)\right]\right| \leq \mathop{\mathbb{E}}_{\vec{X},Y,G}\left[\neg T_{\phi,\vec{X}}(Y \circ G)\right].$$

Observe that $Y \circ G$ is exactly the pseudorandom distribution used to prove Theorem 1.1. Therefore, it ε-fools depth-d read-once $\mathbf{AC}^0$ formulas. Therefore, by Item 1 of Lemma 9.5,
$$\mathop{\mathbb{E}}_{\vec{X},Y,G}\left[\neg T_{\phi,\vec{X}}(Y \circ G)\right] \leq nr(\varepsilon_1 + (s+1)\varepsilon_0).$$

This will be one term in the overall error. Next, we have

$$\begin{aligned}
&\left| \mathop{\mathbb{E}}_{\vec{X},Y,G}\left[\widetilde{\phi}_{\vec{X}}(Y \circ G)\right] - \mathop{\mathbb{E}}_{\vec{X},Y,U}\left[\widetilde{\phi}_{\vec{X}}(Y \circ U)\right]\right| \\
&\qquad \leq \left| \mathop{\mathbb{E}}_{\vec{X}}\left[\mathop{\mathbb{E}}_{Y,G}\left[\left(\widetilde{\phi}_{\vec{X}}|_Y\right)(G)\right] \;\middle|\; E\right] - \mathop{\mathbb{E}}_{\vec{X}}\left[\mathop{\mathbb{E}}_{Y,U}\left[\left(\widetilde{\phi}_{\vec{X}}|_Y\right)(U)\right] \;\middle|\; E\right]\right| + 2\Pr_{\vec{X}}[\neg E] \\
&\qquad \leq \frac{\varepsilon_0}{2} + 2\Pr_{\vec{X}}[\neg E],
\end{aligned}$$

where the last step was by Lemma 9.6 (note that $\Delta\left(\widetilde{\phi}_{\vec{X}}|_Y\right) \leq \Delta\left(\widetilde{\phi}_{\vec{X}}\right)$.) This is another term in the overall error. For the next step, by Lemma 9.2, we have

$$\left| \mathop{\mathbb{E}}_{\vec{X},Y,U}\left[\widetilde{\phi}_{\vec{X}}(Y \circ U)\right] - \mathop{\mathbb{E}}_{\vec{X},U}\left[\widetilde{\phi}_{\vec{X}}(U)\right]\right| \leq s\varepsilon_0.$$

Now, trivially, U fools read-once $\mathbf{AC}^0$ with error 0, so

$$\begin{aligned}
\left| \mathop{\mathbb{E}}_{\vec{X},U}\left[\widetilde{\phi}_{\vec{X}}(U)\right] - \mathop{\mathbb{E}}_{\vec{X},U}\left[\phi|_{X_1 \circ \cdots \circ X_r}(U)\right]\right| &\leq \mathop{\mathbb{E}}_{\vec{X},U}\left[\neg T_{\phi,\vec{X}}(U)\right] && \text{By Item 1 of Lemma 9.5} \\
&\leq nr(s+1)\varepsilon_0 && \text{By Item 2 of Lemma 9.5.}
\end{aligned}$$

Invoking Lemma 9.2 one more time gives

$$\left| \mathop{\mathbb{E}}_{\vec{X},U}\left[\phi|_{X_1 \circ \cdots \circ X_r}(U)\right] - \mathop{\mathbb{E}}_{U}[\phi(U)]\right| \leq s\varepsilon_0.$$

Adding up all the errors by the triangle inequality, we get

$$\begin{aligned}
\left| \mathop{\mathbb{E}}_{H_d^{\circ 2s},G}[\phi(H_d^{\circ 2s} \circ G)] - \mathop{\mathbb{E}}_{U}[\phi(U)]\right| &\leq nr(\varepsilon + (s+1)\varepsilon_0) + \frac{\varepsilon_0}{2} + 2\Pr[\neg E] + s\varepsilon_0 + nr(s+1)\varepsilon_0 + s\varepsilon_0 \\
&\leq nr\varepsilon_1 + 5nrs\varepsilon_0 \\
&< n^2\varepsilon_1
\end{aligned}$$

as claimed. ☐

9.2 The general case of read-once $\mathbf{AC}^0[\oplus]$ with t parity gates

We first prove a seemingly weak bound on the spectral norm (i.e. the sum of the absolute value of the Fourier coefficients) of a read-once $\mathbf{AC}^0[\oplus]$ formula ϕ in terms of the number of its gates, denoted as $\mathrm{size}(\phi)$.

Lemma 9.7. *Let ϕ be an $\mathbf{AC}^0[\oplus]$ formula. Then,*

$$\left\|\widehat{\phi}\right\|_1 \leq 3^{\mathrm{size}(\phi)}.$$

Proof. The proof uses the fact that spectral norm behaves nicely under composition.

Claim 9.8. *Let $f(x) = g(h_1(x), ..., h_m(x))$, where $f: \{0,1\}^n \to \{-1,1\}$, $g: \{-1,1\}^m \to \{-1,1\}$. Then,*

$$\left\|\widehat{f}\right\|_1 \le \|\widehat{g}\|_1 \cdot \prod_{i=1}^{m} \left\|\widehat{h_i}\right\|_1$$

Proof. Note that,

$$f(x) = \sum_{S\subseteq[n]} \widehat{g}(S) \prod_{i\in S} h_i(x).$$

The triangle inequality and submultiplicativity of the spectral norm give

$$\left\|\widehat{f}\right\| \le \sum_{S\subseteq[n]} |\widehat{g}(S)| \prod_{i\in S} \left\|\widehat{h_i}\right\|_1 \le \sum_{S\subseteq[n]} |\widehat{g}(S)| \prod_{i=1}^{m} \left\|\widehat{h_i}\right\|_1 = \|\widehat{g}\|_1 \cdot \prod_{i=1}^{m} \left\|\widehat{h_i}\right\|_1,$$

where the second inequality uses the fact that $\left\|\widehat{h_i}\right\|_1 \ge 1$, as can be seen as follows. Choose an arbitrary $x \in \{0,1\}^n$, we have

$$1 = |h_i(x)| = \left|\sum_{S\subseteq[n]} \widehat{h}_i(S)\chi_S(x)\right| \le \sum_{S\subseteq[n]} \left|\widehat{h}_i(S)\right| = \left\|\widehat{h_i}\right\|_1.$$

□

Let $\wedge_m, \vee_m, \oplus_m : \{0,1\}^m \to \{0,1\}$ denote an $\wedge$ gate with m inputs, an $\vee$ gate with m inputs, and a $\oplus$ gate with m inputs respectively. We use the fact that for any $m > 0$,

$$\left\|\widehat{(-1)^{\wedge_m}}\right\|_1, \left\|\widehat{(-1)^{\vee_m}}\right\|_1, \left\|\widehat{(-1)^{\oplus_m}}\right\|_1 \le 3.$$

Let $\mathcal{G}$ denote the set of the gates in the circuit ϕ. Applying Claim 9.8 recursively over all the gates of ϕ implies that

$$\left\|\widehat{\phi}\right\|_1 \le \frac{1}{2} + \frac{1}{2} \cdot \left\|\widehat{(-1)^{\phi}}\right\|_1 \le \frac{1}{2} + \frac{1}{2} \cdot \prod_{g\in\mathcal{G}} \left\|\widehat{(-1)^{g}}\right\| \le \frac{1}{2} + \frac{1}{2} \cdot 3^{|\mathcal{G}|} \le 3^{\text{size}(\phi)}.$$

□

Proposition 9.9. *Let ϕ be a depth-$(d+1)$ read-once $\mathbf{AC}^0[\oplus]$ formula with $t \ge 1$ parity gates. Then $H_d^{\circ 2s} \circ G_{\text{MRT}}$ fools f with error $n^2\varepsilon_1 \cdot 3^{(d+1)t}$.*

Proof. Let A denote the set of all gates of ϕ that are either a parity gate or have a descendant that is a parity gate. It is easy to see that $|A| \le (d+1)t$, since each parity gate contributes to at most $d+1$ ancestors. Define $Y = \{y_1, ..., y_m\}$ to be the set of all nodes outside A that have an immediate parent in A, moreover, let $h_1, ..., h_m$ to be the functions computed at these nodes respectively. It is easy to see that,

$$\phi(x) = g(h_1, ..., h_m),$$

where g is a depth-d read-once $\mathbf{AC}^0[\oplus]$ formula of size at most $(d+1)t$. Using the Fourier expansion of g,

$$\phi(x) = \sum_{S\subseteq[m]} \widehat{g}(S) \cdot \prod_{i\in S} (-1)^{h_i} = \sum_{S\subseteq[m]} \widehat{g}(S) \cdot (1 - 2 \cdot \bigoplus_{i\in S} h_i).$$

By Lemma 9.7, $\|\widehat{g}\|_1 \le 3^{(d+1)t}$, and by Lemma 9.1, each $\bigoplus_{i\in S} h_i$ is $n^2\varepsilon$ fooled by $H_d^{\circ 2s} \circ G_{\text{MRT}}$. As a result $H_d^{\circ 2s} \circ G_{\text{MRT}}$ fools ϕ with error at most $2 \cdot n^2\varepsilon_1 \cdot 3^{(d+1)t}$. □

Proof of Theorem 1.2. By Proposition 9.9, the generator $H_d^{\circ 2s} \circ G_{\mathrm{MRT}}$ ε-fools depth-$(d+1)$ read-once $\mathbf{AC}^0[\oplus]$ formulas with at most t parity gates provided we set $\varepsilon_1 := \frac{\varepsilon}{n^2 \cdot 3^{(d+1)t}}$. Now we bound the seed length. The seed length for the distributions D and T underlying H_d is still bounded by $O(d^2 \log(n/\varepsilon_1) \log\log(n/\varepsilon_1) \log\log n)$, just as in the proof of Theorem 1.1. Similarly, the seed length for G_d and G_{MRT} is still bounded by

$$\log(n/\varepsilon_1) \cdot O((d+1)\log\log(n/\varepsilon_1))^{2(d+1)+2}$$
$$((d+1)t + \log(n/\varepsilon)) \cdot O((d+1)(\log\log(n/\varepsilon) + \log((d+1)t))^{2(d+1)+2}.$$

This second term dominates. Replacing d with $d-1$ completes the proof. □

10 Acknowledgments

We thank David Zuckerman for very helpful discussions. The first author would also like to thank Gil Cohen, Chin Ho Lee and Amnon Ta-Shma for insightful conversations about the Forbes-Kelley result [FK18].

References

[AGHP92] Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. Simple constructions of almost k-wise independent random variables. *Random Structures & Algorithms*, 3(3):289–304, 1992.

[AW89] Miklos Ajtai and Avi Wigderson. Deterministic simulation of probabilistic constant depth circuits. *Advances in Computing Research*, 5(199-222):1, 1989.

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

[BDVY13] Andrej Bogdanov, Zeev Dvir, Elad Verbin, and Amir Yehudayoff. Pseudorandomness for width-2 branching programs. *Theory of Computing*, 9:283–292, 2013.

[BPW11] Andrej Bogdanov, Periklis A Papakonstaninou, and Andrew Wan. Pseudorandomness for read-once formulas. In *Proceedings of the 52nd Annual IEEE Symposium on Foundations of Computer Science (FOCS 2011)*, pages 240–246. IEEE, 2011.

[Bra09] Mark Braverman. Poly-logarithmic independence fools $\mathbf{AC}^0$ circuits. In *Proceedings of the 24th Annual IEEE Conference on Computational Complexity (CCC 2009)*, pages 3–8. IEEE, 2009.

[BRRY14] Mark Braverman, Anup Rao, Ran Raz, and Amir Yehudayoff. Pseudorandom generators for regular branching programs. *SIAM Journal on Computing*, 43(3):973–986, 2014.

[CH05] Arkadev Chattopadhyay and Kristoffer Arnsfelt Hansen. Lower bounds for circuits with few modular and symmetric gates. In *Automata, languages and programming*, volume 3580 of *Lecture Notes in Comput. Sci.*, pages 994–1005. Springer, Berlin, 2005.

[CHHL18] Eshan Chattopadhyay, Pooya Hatami, Kaave Hosseini, and Shachar Lovett. Pseudorandom generators from polarizing random walks. In *Proceedings of the 33rd Annual Computational Complexity Conference (CCC 2018)*, volume 102 of *LIPIcs. Leibniz Int.*

Proc. Inform., pages Art. No. 1, 21. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2018.

[CHRT18] Eshan Chattopadhyay, Pooya Hatami, Omer Reingold, and Avishay Tal. Improved pseudorandomness for unordered branching programs through local monotonicity. In *Proceedings of the 50th Annual ACM Symposium on Theory of Computing (STOC 2018)*, pages 363–375, New York, NY, USA, 2018. ACM.

[CRS00] Suresh Chari, Pankaj Rohatgi, and Aravind Srinivasan. Improved algorithms via approximations of probability distributions. *J. Comput. System Sci.*, 61(1):81–107, 2000.

[CSV15] Sitan Chen, Thomas Steinke, and Salil Vadhan. Pseudorandomness for read-once, constant-depth circuits. *arXiv preprint arXiv:1504.04675*, 2015.

[CZ16] Eshan Chattopadhyay and David Zuckerman. Explicit two-source extractors and resilient functions. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing*, pages 670–683. ACM, 2016.

[De11] Anindya De. Pseudorandomness for permutation and regular branching programs. In *Proceedings of the 26th Annual IEEE 26th Annual Conference on Computational Complexity (CCC 2011)*, pages 221–231. IEEE, 2011.

[DETT10] Anindya De, Omid Etesami, Luca Trevisan, and Madhur Tulsiani. Improved pseudorandom generators for depth 2 circuits. In *Approximation, randomization, and combinatorial optimization*, volume 6302 of *Lecture Notes in Comput. Sci.*, pages 504–517. Springer, Berlin, 2010.

[FK18] Michael A. Forbes and Zander Kelley. Pseudorandom generators for read-once branching programs, in any order. In *Proceedings of the 59th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2018)*. IEEE, 2018.

[FSS84] Merrick Furst, James B Saxe, and Michael Sipser. Parity, circuits, and the polynomial-time hierarchy. *Mathematical systems theory*, 17(1):13–27, 1984.

[FSUV13] Bill Fefferman, Ronen Shaltiel, Christopher Umans, and Emanuele Viola. On beating the hybrid argument. *Theory Comput.*, 9:809–843, 2013.

[GLS12] Dmitry Gavinsky, Shachar Lovett, and Srikanth Srinivasan. Pseudorandom generators for read-once $\mathbf{ACC}^0$. In *Proceedings of the 27th Annual IEEE Conference on Computational Complexity (CCC 2012)*, pages 287–297, 2012.

[GMR+12] Parikshit Gopalan, Raghu Meka, Omer Reingold, Luca Trevisan, and Salil Vadhan. Better pseudorandom generators from milder pseudorandom restrictions. In *Proceedings of the 53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS 2012)*, pages 120–129. IEEE, 2012.

[GMR13] Parikshit Gopalan, Raghu Meka, and Omer Reingold. DNF sparsification and a faster deterministic counting algorithm. *Comput. Complexity*, 22(2):275–310, 2013.

[GW14] Oded Goldreich and Avi Widgerson. On derandomizing algorithms that err extremely rarely. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing (STOC 2014)*, pages 109–118. ACM, New York, 2014.

[Has86] Johan Haståd. Almost optimal lower bounds for small depth circuits. In *Proceedings of the eighteenth Annual ACM Symposium on Theory of Computing (STOC 1986)*, pages 6–20. ACM, 1986.

[HLV18] Elad Haramaty, Chin Ho Lee, and Emanuele Viola. Bounded independence plus noise fools products. *SIAM J. Comput.*, 47(2):493–523, 2018.

[HS16] Prahladh Harsha and Srikanth Srinivasan. On polynomial approximations to $\mathbf{AC}^0$. *arXiv preprint arXiv:1604.08121*, 2016.

[IMZ12] Russell Impagliazzo, Raghu Meka, and David Zuckerman. Pseudorandomness from shrinkage. In *Proceedings of the 53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS 2012)*, pages 111–119. IEEE, 2012.

[KLW10] Adam R. Klivans, Homin Lee, and Andrew Wan. Mansour's conjecture is true for random DNF formulas. In *Proceedings of the 23rd Annual Conference on Learning Theory (COLT 2010)*, 2010.

[KNP11] Michal Koucký, Prajakta Nimbhorkar, and Pavel Pudlák. Pseudorandom generators for group products. In *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing (STOC 2011)*, pages 263–272. ACM, New York, 2011.

[LN90] Nathan Linial and Noam Nisan. Approximate inclusion-exclusion. *Combinatorica*, 10(4):349–365, 1990.

[LS11] Shachar Lovett and Srikanth Srinivasan. Correlation bounds for poly-size $\mathbf{AC}^0$ 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.

[LVW93] Michael Luby, Boban Velickovic, and Avi Wigderson. Deterministic approximate counting of depth-2 circuits. In *Proceedings of the 2nd Annual Israel Symposium on Theory and Computing Systems (ISTCS 1993)*, pages 18–24. IEEE, 1993.

[MRT19] Raghu Meka, Omer Reingold, and Avishay Tal. Pseudorandom generators for width-3 branching programs. In *Proceedings of the 51st Annual ACM Symposium on Theory of Computing (STOC 2019)*, 2019. To appear.

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

[Nis92] Noam Nisan. Pseudorandom generators for space-bounded computation. *Combinatorica*, 12(4):449–461, 1992.

[NN93] Joseph Naor and Moni Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM Journal on Computing*, 22(4):838–856, 1993.

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

[RSV13] Omer Reingold, Thomas Steinke, and Salil Vadhan. Pseudorandomness for regular branching programs via Fourier analysis. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 655–670. Springer, 2013.

[ST18] Rocco A. Servedio and Li-Yang Tan. Improved pseudorandom generators from pseudorandom multi-switching lemmas. *arXiv preprint arXiv:1801.03590*, 2018.

[ST19] Rocco A. Servedio and Li-Yang Tan. Pseudorandomness for read-k DNF formulas. In *Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2019)*, pages 621–638, 2019.

[Ste12] Thomas Steinke. Pseudorandomness for permutation branching programs without the group theory. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 19, page 6, 2012.

[ŠŽ11] Jiří Šíma and Stanislav Žák. Almost k-wise independent sets establish hitting sets for width-3 1-branching programs. In *Computer science—theory and applications*, volume 6651 of *Lecture Notes in Comput. Sci.*, pages 120–133. Springer, Heidelberg, 2011.

[Tal17] Avishay Tal. Tight Bounds on the Fourier Spectrum of $\mathbf{AC}^0$. In Ryan O'Donnell, editor, *Proceedings of the 32nd Annual Computational Complexity Conference (CCC 2017)*, volume 79 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 15:1–15:31, Dagstuhl, Germany, 2017. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[TX13] Luca Trevisan and Tongke Xue. A derandomized switching lemma and an improved derandomization of $\mathbf{AC}^0$. In *Proceedings of the 28th Annual IEEE Conference on Computational Complexity (CCC 2013)*, pages 242–247. IEEE, 2013.

[Vio07] Emanuele Viola. Pseudorandom bits for constant-depth circuits with few arbitrary symmetric gates. *SIAM J. Comput.*, 36(5):1387–1403, 2006/07.

A Proofs of Lemma 6.1 and Lemma 9.4

Recall that Lemma 6.1 states that every read-once NAND formula can be sandwiched by two similar structured NAND formulas where every gate has a non-negligible chance of rejecting. We now present the proof of Lemma 6.1. We emphasize that this argument was already given by Chen, Steinke, and Vadhan [CSV15]; we are reproducing it here to verify the exact parameters of Lemma 6.1 and so that we can reference the proof when proving Lemma 9.4.

Proof of Lemma 6.1. We proceed by induction on $\mathrm{size}(\phi)$, i.e., the number of NAND gates, to prove the lemma with the modified bound $\mathbb{E}[u_\phi - \ell_\phi] \leq n\sqrt{\theta} + \mathrm{size}(\phi)\theta$. In the base case $\mathrm{size}(\phi) = 0$, if ϕ is non-constant, it is a single literal, which has expectation $\frac{1}{2}$, so we can simply take $\ell_\phi = u_\phi = \phi$. Now for the inductive step, suppose $\phi = \mathrm{NAND}(\phi_1, \dots, \phi_m)$. Let n_i be the number of inputs to ϕ_i, so $\sum_i n_i = n$ (recall ϕ is read-once). By induction, for each $i \in [m]$, there exist formulas $\ell_{\phi_i} \leq \phi_i \leq u_{\phi_i}$ with the following properties:

- $\mathbb{E}[u_{\phi_i} - \ell_{\phi_i}] \leq n_i\sqrt{\theta} + \mathrm{size}(\phi_i)\theta$.
- Each of u_{ϕ_i} and ℓ_{ϕ_i} has an underlying tree structure that is a subgraph of the underlying tree structure of ϕ_i.
- Every non-constant gate ψ in either ℓ_{ϕ_i} or u_{ϕ_i} satisfies $\mathbb{E}[\psi] \geq \theta$ and $\mathbb{E}[\neg\psi] \geq \theta$.

We consider two cases. For the first case, suppose $\mathbb{E}[\neg\phi] \geq \theta$. In this case, define

$$\begin{aligned}
\ell_\phi &= \mathrm{NAND}(u_{\phi_1}, \ldots, u_{\phi_m}) \\
u_\phi &= \begin{cases} \mathrm{NAND}(\ell_{\phi_1}, \ldots, \ell_{\phi_m}) & \text{if that gives } \mathbb{E}[\neg u_\phi] \geq \theta \\ 1 & \text{otherwise.} \end{cases}
\end{aligned}$$

Because NAND is anti-monotone, $\ell_\phi \leq \phi \leq u_\phi$. In the first case of the definition of u_ϕ, by the union bound, we have

$$\mathbb{E}[u_\phi - \ell_\phi] \leq \sum_{i=1}^{m}(n_i\sqrt{\theta} + \mathrm{size}(\phi_i)\theta) = n\sqrt{\theta} + (\mathrm{size}(\phi) - 1)\theta$$

as desired. In the second case of the definition of u_ϕ, the error only increases by at most θ, which is still within the bound of $n\sqrt{\theta} + \mathrm{size}(\phi)\theta$. Finally, we must verify that every non-constant gate ψ in these formulas satisfies $\mathbb{E}[\psi] \geq \theta$ and $\mathbb{E}[\neg\psi] \geq \theta$. For gates other than the output gate, this is true by induction, so let us verify that it holds for the output gates. We have $\mathbb{E}[\neg\ell_\phi] \geq \mathbb{E}[\neg\phi] \geq \theta$. On the other hand, if ℓ_ϕ is non-constant, then some child u_{ϕ_i} is non-constant, hence $\mathbb{E}[\ell_\phi] \geq \mathbb{E}[\neg u_{\phi_i}] \geq \theta$. Similarly, by construction, if u_ϕ is non-constant, then $\mathbb{E}[\neg u_\phi] \geq \theta$ and $\mathbb{E}[u_\phi] \geq \mathbb{E}[\neg\ell_{\phi_i}] \geq \theta$.

Now, for the second case, suppose $\mathbb{E}[\neg\phi] < \theta$. In this case, define

$$\begin{aligned}
\widetilde{\ell}_\phi &= \mathrm{NAND}(u_{\phi_1}, \ldots, u_{\phi_m}) \\
u_\phi &= 1.
\end{aligned}$$

As before, $\widetilde{\ell}_\phi \leq \phi \leq u_\phi$, and if $\widetilde{\ell}_\phi$ is non-constant, then $\mathbb{E}[\widetilde{\ell}_\phi] \geq \mathbb{E}[\neg u_{\phi_i}] \geq \theta$. Furthermore, $\mathbb{E}[u_\phi - \widetilde{\ell}_\phi] \leq n\sqrt{\theta} + \mathrm{size}(\phi)\theta$. So if $\mathbb{E}[\neg\widetilde{\ell}_\phi] \geq \theta$, we can just set $\ell_\phi = \widetilde{\ell}_\phi$ and we're done. Assume, therefore, that $\mathbb{E}[\neg\widetilde{\ell}_\phi] < \theta$.

In this case, we divide into two subcases. First, suppose that for some i, we have $\mathbb{E}[u_{\phi_i}] \leq \sqrt{\theta}$. Then we define $\ell_\phi = \mathrm{NAND}(u_{\phi_i})$. Clearly, we still have $\ell_\phi \leq \phi$. Furthermore,

$$\mathbb{E}[u_\phi - \ell_\phi] = \mathbb{E}[\neg\ell_\phi] = \mathbb{E}[u_{\phi_i}] \leq \sqrt{\theta}.$$

For the second and final subcase, suppose that for every i, $\mathbb{E}[u_{\phi_i}] > \sqrt{\theta}$. In this case, since $\prod_{i=1}^{m}\mathbb{E}[u_{\phi_i}] = \mathbb{E}[\neg\widetilde{\ell}_\phi] < \theta$, there must be some j such that

$$\theta \leq \prod_{i=1}^{j}\mathbb{E}[u_{\phi_i}] \leq \sqrt{\theta}.$$

Therefore, define

$$\ell_\phi = \mathrm{NAND}(u_{\phi_1}, \ldots, u_{\phi_j}).$$

That way, $\ell_\phi \leq \phi \leq u_\phi$, and $\mathbb{E}[\neg\ell_\phi] \geq \theta$, and

$$\mathbb{E}[u_\phi - \ell_\phi] = \mathbb{E}[\neg\ell_\phi] \leq \sqrt{\theta}.$$

That completes the induction. To get the parameters claimed in the lemma statement, just observe that $\mathrm{size}(\phi) \leq nd$ and $n\sqrt{\theta} + nd\theta < \varepsilon_0$. $\square$

Now we state and prove a strengthening of Lemma 9.4.

Lemma A.1. *Let ϕ be a depth-d read-once* NAND *formula over n variables with $d \le n$ and let $\varepsilon_0 > 0$. Let ℓ_ϕ and u_ϕ be the read-once* NAND *formulas guaranteed to us by Lemma 6.1. Then, there exist $T^\ell_\phi, T^u_\phi \colon \{0,1\}^n \to \{0,1\}$ satisfying the following conditions:*

1. *If $x \in \{0,1\}^n$ is such that $\phi(x) \neq \ell_\phi(x)$ then $T^\ell_\phi(x) = 0$.*
2. *If $x \in \{0,1\}^n$ is such that $\phi(x) \neq u_\phi(x)$ then $T^u_\phi(x) = 0$.*
3. *Both $\mathbb{E}[T^\ell_\phi] \ge 1 - \varepsilon_0$ and $\mathbb{E}[T^\ell_\phi] \ge 1 - \varepsilon_0$.*
4. *Both T^ℓ_ϕ and T^u_ϕ are computable by depth-d read-once $\mathbf{AC}^0$ formulas.*

Roughly speaking, the lemma gives us an "error-indicator" read-once formula that is guaranteed to be zero whenever the sandwiching formula does not give the same value as the original formula. The proof of the lemma will heavily use the proof of Lemma 6.1.

Proof. The proof is by induction on $\mathrm{size}(\phi)$, as in Lemma 6.1. In the base case $\mathrm{size}(\phi) = 0$, we simply take $T^\ell_\phi = T^u_\phi = 1$ since $\ell_\phi = u_\phi = \phi$. For the inductive step, suppose $\phi = \mathrm{NAND}(\phi_1, \ldots, \phi_m)$ where for each i, $\mathrm{size}(\phi_i) = n_i$ so that $\sum_i n_i = n$. By our hypothesis, for every $i \in [m]$ there exist formulas ℓ_{ϕ_i} and u_{ϕ_i} guaranteed to us by Lemma 6.1, as well as formulas $T^u_{\phi_i}$ and $T^u_{\phi_i}$ with the following properties:

- $T^\ell_{\phi_i}(x) = 0$ whenever $\phi_i(x) \neq \ell_{\phi_i}$.
- $T^u_{\phi_i}(x) = 0$ whenever $\phi_i(x) \neq u_{\phi_i}$.
- $\mathbb{E}[\neg T^\ell_{\phi_i}] \le n_i\sqrt{\theta} + \mathrm{size}(\phi_i)\theta$ and $\mathbb{E}[\neg T^u_{\phi_i}] \le n_i\sqrt{\theta} + \mathrm{size}(\phi_i)\theta$, for $\theta = \frac{\varepsilon_0^2}{4n^2}$.
- $T^\ell_{\phi_i}$ and $T^u_{\phi_i}$ are computable by depth-$(d-1)$ read-once $\mathbf{AC}^0$ formulas.

Let us first handle T^u_ϕ. For u_ϕ there are two possibilities. It can be either set to $u_\phi = 1$ or set to $u_\phi = \mathrm{NAND}(\ell_{\phi_1}, \ldots, \ell_{\phi_m})$.

1. In the first case, where $u_\phi = 1$, we set $T^u_\phi = \phi$ and so when $T^u_\phi(x) = 1$ clearly $\phi(x) = u_\phi(x) = 1$. To bound $\mathbb{E}[T^u_\phi] = \mathbb{E}[\phi]$, recall that this case is invoked only when either $\mathbb{E}[\neg\phi] < \theta$, in which case the bound is clear, or when $\mathbb{E}[\ell_{\phi_1} \wedge \ldots \wedge \ell_{\phi_m}] = \prod_i \mathbb{E}[\ell_{\phi_i}] < \theta$. In the latter case, since $\mathbb{E}[\phi_i - \ell_{\phi_i}] \le n_i\sqrt{\theta} + \mathrm{size}(\phi_i)\theta \triangleq \zeta_i$, we obtain
$$\begin{aligned}\mathbb{E}[\neg T^u_\phi] &= \prod_{i=1}^m \mathbb{E}[\phi_i] \\ &= \prod_{i=1}^m \mathbb{E}[\ell_{\phi_i}] + \sum_{i=1}^m \left((\mathbb{E}[\phi_i] - \mathbb{E}[\ell_{\phi_i}]) \prod_{j=1}^{i-1} \mathbb{E}[\phi_j] \prod_{j=i+1}^{m} \mathbb{E}[\ell_{\phi_j}] \right) \\ &\le \theta + \sum_{i=1}^m \zeta_i = \theta + n\sqrt{\theta} + (\mathrm{size}(\phi) - 1)\theta = n\sqrt{\theta} + \mathrm{size}(\phi)\theta \le \varepsilon_0.\end{aligned}$$

2. In the second case, where $u_\phi = \mathrm{NAND}(\ell_{\phi_1}, \ldots, \ell_{\phi_m})$, set $T^u_\phi = \bigwedge_{i=1}^m T^\ell_{\phi_i}$. If $x \in \{0,1\}^n$ is such that $T^u_\phi(x) = 1$ then $T^\ell_{\phi_i}(x) = 1$ for every $i \in [m]$ and so $u_\phi(x) = \mathrm{NAND}(\phi_1(x), \ldots, \phi_m(x)) = \phi(x)$. To bound $\mathbb{E}[T^u_\phi]$, note that
$$\Pr[T^u_\phi = 0] \le \sum_{i=1}^m \Pr[T^\ell_{\phi_i} = 0] \le \sum_{i=1}^m (n_i\sqrt{\theta} + \mathrm{size}(\phi_i)\theta) = n\sqrt{\theta} + (\mathrm{size}(\phi) - 1)\theta \le \varepsilon_0,$$

as desired.

In both cases, the depth requirement is immediate.

We shall now handle T^ℓ_ϕ. The two possibilities for ℓ_ϕ are as follows.

1. In the first case, $\ell_\phi = \mathrm{NAND}(u_{\phi_1}, \ldots, u_{\phi_m})$. Here, we set $T^\ell_\phi = \bigwedge_{i=1}^m T^u_{\phi_i}$ and the correctness is similar to case (2) of T^u_ϕ.

2. In the second case, up to reordering of the formulas, there exists $j \in [m-1]$ such that $\ell_\phi = \mathrm{NAND}(u_{\phi_1}, \ldots, u_{\phi_j})$. We choose $T^\ell_\phi = \ell_\phi$, and surely if $x \in \{0,1\}^n$ is such that $\ell_\phi(x) = 1$ then $\phi(x) = 1$ since $\phi \geq \ell_\phi$.

   To bound $\mathbb{E}[T^\ell_\phi]$, recall that j is chosen (again, up to reordering) so that $\mathbb{E}[u_{\phi_1} \wedge \ldots \wedge u_{\phi_j}] \leq \sqrt{\theta}$. Thus, $\mathbb{E}[\neg T^\ell_\phi] \leq \sqrt{\theta} \leq \varepsilon_0$.

Again, in both cases, the depth requirement is immediate. □

B Proof of Lemma 9.5

Toward proving Lemma 9.5, fix ϕ, define $X_0 = \star^n$, and define $\ell^{(0)}_{j,\vec{X}} = \phi_j$ for each $j \in [m]$. Then, for $i < r$, inductively define

$$\ell^{(i+1)}_{j,\vec{X}} = \ell_{(\ell^{(i)}_{j,\vec{X}}|_{X_i})}$$

That is, $\ell^{(i+1)}_{j,\vec{X}}$ is the lower sandwiching formula when Lemma 6.1 is applied to $\ell^{(i)}_{j,\vec{X}}|_{X_i}$. Furthermore, define

$$T^{(i+1)}_{j,\vec{X}} = T^\ell_{(\ell^{(i)}_{j,\vec{X}}|_{X_i})}|_{X_{i+1}\circ\cdots\circ X_r}.$$

That is, $T^{(i+1)}_{j,\vec{X}}$ is the success certifier of Lemma 9.4 for the sandwiching formula $\ell^{(i+1)}_{j,\vec{X}}$, restricted according to $X_{i+1} \circ \cdots \circ X_r$. Finally, define

$$\widetilde{\phi}_{\vec{X}} = \bigoplus_{j=1}^m \left(\ell^{(r)}_{j,\vec{X}}|_{X_r}\right)$$

$$T_{\phi,\vec{X}} = \bigwedge_{i=1}^r \bigwedge_{j=1}^m T^{(i)}_{j,\vec{X}}.$$

Proof of Item 1 of Lemma 9.5. Fix x and assume $T_{\phi,\vec{X}}(x) = 1$. Fix an arbitrary $j \in [m]$. We'll show by backward induction on i that

$$(\ell^{(r)}_{j,\vec{X}}|_{X_r})(x) = (\ell^{(i)}_{j,\vec{X}}|_{X_i\circ X_{i+1}\circ\cdots\circ X_r})(x). \tag{7}$$

In the base case $i = r$, this is trivial. Now for the inductive step, assume Equation (7) is true for $i+1$, and we'll prove it for i. Since $T_{\phi,\vec{X}}(x) = 1$, we must have $T^{(i+1)}_{\phi,\vec{X}}(x) = 1$. That is, $(T^\ell_{(\ell^{(i)}_{j,\vec{X}}|_{X_i})}|_{X_{i+1}\circ\cdots\circ X_r})(x) = 1$. This implies that

$$\ell^{(i+1)}_{j,\vec{X}}(X_{i+1} \circ \cdots \circ X_r \circ x) = \ell^{(i)}_{j,\vec{X}}(X_i \circ X_{i+1} \circ \cdots \circ X_r \circ x).$$

Applying the induction hypothesis completes the proof of Equation (7). Now, plugging in $i = 0$ to Equation (7), we find that

$$(\ell^{(r)}_{j,\vec{X}}|_{X_r})(x) = (\phi_j|_{X_1 \circ \cdots \circ X_r})(x). \tag{8}$$

Since Equation (8) holds for all j simultaneously, we can apply the parity operation from $j = 1$ to m to complete the proof. □

Proof of Item 2 of Lemma 9.5. Fix any arbitrary $i \in [r], j \in [m]$. Let $U \sim U_n$ be independent of $\vec{X}$. We have

$$\begin{aligned}
\mathop{\mathbb{E}}_{\vec{X},G}\left[T^{(i)}_{j,\vec{X}}(G)\right] &\geq \mathop{\mathbb{E}}_{\vec{X},U}\left[T^{(i)}_{j,\vec{X}}(U)\right] - \varepsilon_1 && \text{Because } T^{(i)}_{j,\vec{X}} \text{ is a depth-}d\text{ formula} \\
&= \mathop{\mathbb{E}}_{\vec{X},U}\left[T^{\ell}_{(\ell^{(i-1)}_{j,\vec{X}}|_{X_{i-1}})}(X_i \circ \cdots \circ X_r \circ U)\right] - \varepsilon_1 && \text{By the definition of } T^{(i)}_{j,\vec{X}} \\
&\geq \mathop{\mathbb{E}}_{\vec{X},U}\left[T^{\ell}_{(\ell^{(i-1)}_{j,\vec{X}}|_{X_{i-1}})}(U)\right] - \varepsilon_1 - s\varepsilon_0 && \text{By Corollary 4.3} \\
&\geq 1 - \varepsilon_1 - (s+1)\varepsilon_0 && \text{By Lemma 9.4.}
\end{aligned}$$

Taking a union bound over i and j completes the proof. □

The proof of Item 3 of Lemma 9.5 is essentially the same as the proof of Item 3 of Lemma 6.2 and we omit it.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il