Electronic Colloquium on Computational Complexity, 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 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]. Our result improves on the Forbes-Kelley PRG even when d is slightly super-constant.

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 formulas. To fool depth-$(d+1)$ 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 $\text{poly}(\log \log n)$ independent applications of our pseudorandom restriction, the formula simplifies in the sense that every gate other than the output has only $\text{polylog}\, n$ remaining children. Finally, as the last step, we use a recent PRG by Meka, Reingold, and Tal [MRT18] 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

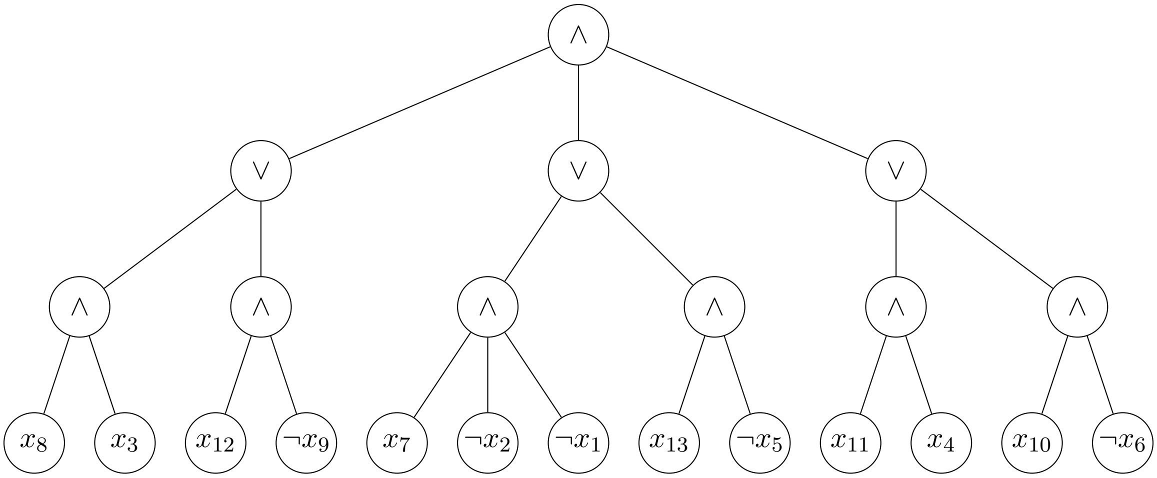

Figure 1: A depth-3 read-once formula on $n = 13$ bits.

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). 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 formulas with seed length*

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

1.1 Motivation and prior work

Derandomizing small-space algorithms. We are motivated by the **L** vs. **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 **L** vs. **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, MRT18].

Surprisingly, the study of fooling constant-width ROBPs has so far been closely entangled with the study of fooling constant-depth read-once formulas. A depth-d read-once 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.

Meanwhile, for any constant d, Chen, Steinke and Vadhan constructed a PRG for depth-d read-once 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 $O(\log^3 n)$. In the constant-width case, their PRG has seed length $\widetilde{O}(\log^2 n)$; prior to the present work, this was also the best PRG for constant-depth read-once formulas. 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 formulas, we hope that our result will serve as a stepping stone toward optimal PRGs for general constant-width ROBPs.

Fooling constant-depth circuits. The model we study in this paper (constant-depth read-once formulas) is the read-once version of $\mathbf{AC}^0$. 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 is Trevisan and Xue's PRG with seed length $\widetilde{O}(\log^{d+4} n)$ [TX13]. 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

[1] Note that Nisan's generator [Nis92] is not guaranteed to fool read-once formulas because of the issue of variable ordering [BPW11].

the error ε are constant, their PRG has seed length $2^{O(d^2)} \cdot \log^{O(d)} n$; this result is also subsumed 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]. Klivans, Lee, and Wan extended this result to read-k CNFs for any constant k [KLW10]. As mentioned previously, Gopalan et al. gave a PRG for read-once CNFs with seed length $\widetilde{O}(\log(n/\varepsilon))$ [GMR$^+$12].

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]

Forbes and Kelley constructed a very simple pseudorandom distribution over restrictions that approximately preserves the expectation of any constant-width ROBP [FK18], hence any constant-depth read-once 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, we can simply appeal to the Forbes-Kelley result [FK18] to argue that the expectation of the formula is preserved under our pseudorandom restriction.

[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* constant-depth read-once 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 simply 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 constant-depth read-once 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 constant-depth read-once 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 constant-depth read-once formulas. (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 constant-depth read-once formulas, but the whole reason we are interested in simplifying formulas is to *design* an improved PRG for constant-depth read-once formulas! 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 formulas, and we use G_d to sample pseudorandom restrictions that simplify depth-$(d+1)$ read-once formulas. (This is similar to the approach of Gavinsky et al. [GLS12].) To make this idea work, we overcome several challenges.

- Whether a *single* node is still alive after a random restriction is not the simplification condition we are actually interested in. We are interested in the *number* of remaining living children of each node. To address this issue, we consider the condition that a *collection* of nodes all remain alive, and we relate this condition to the number of living children using an argument introduced by Gopalan et al. in the context of fooling read-once CNFs [GMR+12]. This same argument was also used by Chen et al. in their analysis of truly random restrictions [CSV15].
- To ensure that the Forbes-Kelley analysis [FK18] applies to our scenario, we are forced to design our pseudorandom restriction so that each coordinate is $\star$ with constant probability.

[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 minor technicality is that this is only true "up to sandwiching."

However, in their analysis of truly random restrictions, Chen et al. only showed that formulas simplify when the $\star$-probability is $1/\operatorname{polylog}(n/\varepsilon)$. We therefore have no hope of showing that *one* application of our pseudorandom restriction simplifies formulas. We overcome this difficulty using an approach similar to the method of conditional probabilities. We show that after applying our pseudorandom restriction, the formula is likely to simplify under *additional* truly random restrictions where the *cumulative* $\star$-probability is $1/\operatorname{polylog}(n/\varepsilon)$. Therefore, after applying $O(\log\log(n/\varepsilon))$ independent pseudorandom restrictions, the formula simplifies.

- For a collection of nodes that form subformulas of depth d', we are only able to test the predicate that they are all still alive by a formula of depth $d' + 1$.[5] Therefore, we can only apply G_d when $d' \leq d - 1$. Correspondingly, we are only able to show that after applying $\operatorname{poly}(\log\log(n/\varepsilon))$ independent pseudorandom restrictions, every gate *other than the root* in a depth-$(d+1)$ formula has at most $\operatorname{polylog}(n/\varepsilon)$ living children.[6] Fortunately, the latter condition is strong enough that the restricted formula is fooled by a recent PRG by Meka, Reingold, and Tal [MRT18]. We use the MRT PRG [MRT18] as the last step in our PRG.

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 $f \in \mathcal{C}$,

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

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

A *formula* on $\{0,1\}^n$ is a rooted tree in which each internal node ("gate") is labeled either $\wedge$ 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* formula as one in which the gates are arranged in alternating layers of $\wedge$ and $\vee$ gates. Any read-once formula can be simulated by a layered read-once 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

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

[5] Actually, we don't even quite show that. See Claim 5.6 for the precise statement.

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

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 $\Pr[X_i = 0] = \Pr[X_i = 1] = 1/4$. In general, if H is a distribution over $\{0,1,\star\}^n$ and s is a nonnegative integer, we define $H^{\circ s}$ to be the distribution over $X \in \{0,1,\star\}^n$ obtained by drawing s independent samples $X_1, X_2, \dots, X_s \sim H$ and composing them, $X = X_1 \circ X_2 \circ \dots \circ X_s$. 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

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

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 read-once 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 formulas, and let $\varepsilon > 0$ be arbitrary. We will show how to ε-fool depth-$(d+1)$ read-once 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 [MRT18]. We rely on a PRG by Meka, Reingold, and Tal for XORs of short ROBPs [MRT18]; 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 formulas 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-$(d+1)$ read-once 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,*

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

Ultimately, we are interested in fooling formulas over the basis $\{\wedge, \vee, \neg\}$, 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*

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

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 $t \geq 1$,*

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

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 $\mathbb{E}[\neg\varphi] \leq \rho$ or $\mathbb{E}[\varphi] \leq \rho$ for some $\rho \leq \frac{1}{2}$. Then, for every $\theta \in (0, \frac{2}{n})$ and $p \leq \frac{1}{(9\log(2 \cdot 4^d n/\theta))^d}$ it holds that*

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

To make use of Lemma 5.2, we shall now define a related, intermediate progress measure in addition to Δ. As outlined in Section 1.2, this second progress measure will help us deal with the fact that the parameter p in Lemma 5.2 is only $1/\operatorname{polylog} n$ whereas each coordinate in H_d is $\star$ with probability roughly $1/2$.

Definition 5.3. *Let Φ be a set of formulas over n variables. For an integer $t \geq 1$, we define*

$$\mathrm{fail}_t(\Phi) = \Pr_{X \sim R_n^{\circ t}} [\forall \phi \in \Phi,\ \phi|_X \not\equiv 1].$$

We use Lemma 5.2 to prove:

Lemma 5.4. *Let $\Phi = \{\phi_1, \ldots, \phi_k\}$ be a set of read-once* NAND *formulas over n variables, each of depth $d \leq \log n$ and over disjoint subsets of n variables. Further, assume that for every $i \in [k]$, $\mathbb{E}[\neg\phi_i] \leq \rho$ for some $\rho \leq \frac{1}{2}$. Then, there exists a constant c such that for every $\theta \in (0, \frac{2}{n})$ and integer $t \geq cd \log\log(n/\theta)$,*

$$\mathrm{fail}_t(\Phi) \leq (2\rho + \theta)^k.$$

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

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

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

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

Now,

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

so by the union bound

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

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. □

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 with respect to the $\text{fail}_t(\cdot)$ measure of progress, our pseudorandom restriction H_d simplifies depth-$(d-1)$ formulas just as well as a truly random restriction up to an additive error.

Lemma 5.5. *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 $t \geq 1$,*

$$\mathop{\mathbb{E}}_{X \sim H_d} [\text{fail}_{t-1}(\Phi|_X)] \leq \text{fail}_t(\Phi) + 2\alpha,$$

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 the composition of $t-1$ truly random restrictions that will be applied in the future. 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,

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

Claim 5.6. *For every $d \geq 2$, T_v can be computed by a depth-d read-once formula.*

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

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

2. The test $\phi|_{\text{Res}(y,z)\circ v} \not\equiv 0$ can be computed by a depth-d read-once 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 $\mathrm{NAND}(\ell_1, \dots, \ell_m) \not\equiv 1$ can be tested by a depth-2 formula where the top gate is an $\wedge$, and the predicate $\mathrm{NAND}(\ell_1, \dots, \ell_m) \not\equiv 0$ 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 $\phi = \mathrm{NAND}(\varphi_1, \dots, \varphi_m)$ be a read-once formula of depth d, so each φ_i is a depth-$(d-1)$ read-once formula. We already mentioned that

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

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 formula with a top $\vee$ gate, so overall we get a depth-$(d+1)$ read-once formula with a top $\wedge$ gate. Similarly,

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

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 formula with a top $\wedge$ gate, so overall we get a depth-$(d+1)$ read-once formula with a top $\vee$ gate. □

Recall from Section 3 the distribution

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

We shall later show:

Claim 5.7. *$\overline{G}_d$ (2α)-fools depth-d read-once formulas over $\{0,1\}^{2n}$.*

With the above claim in mind, and Claim 5.6, 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 5.7.

Proof of Claim 5.7. We start by noting that since the class of depth-d read-once formulas is closed under shifts, $G_d\oplus T$ and $G'_d\oplus D$ both α-fool depth-d read-once formulas.

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

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

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

□

Iterating H_d for t times, we get:

Lemma 5.8. *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 $t\geq 1$,*

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

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

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

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

Thus,

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

The third transition used the induction's hypothesis and the last one is due to Lemma 5.8. □

Combining Lemma 5.8 with Lemma 5.4 we immediately get the following corollary.

Corollary 5.9. *Let $\Phi = \{\phi_1, \dots, \phi_k\}$ 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 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 5.10. *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 5.9. 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 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 restriction by a sandwiching argument. The proof of Lemma 5.10 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 $\Psi_i = \{\varphi_1, \dots, \varphi_w\}$. 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 $X \sim H_d^{\circ t}$. Let

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

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

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

We know that $\mathbb{E}[\neg\varphi] \le 2^{i+1}\theta$ and φ is a depth-$(d-1)$ NAND formula, so by Corollary 5.9 we get

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

Now,

Claim 5.11. *It holds that* $w \le \frac{\ln(1/\theta)}{2^i\theta}$.

Proof. On the one hand,

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

On the other hand,

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

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. □

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 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. *$\ell_\phi \le \phi \le u_\phi$ and $\mathbb{E}[u_\phi - \ell_\phi] \le \varepsilon_0$.*

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 $\mathbb{E}[\psi] \ge \theta$ and $\mathbb{E}[\neg\psi] \ge \theta$.*

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 5.10, 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 5.10. 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 $d \le \log n$ and let $\varepsilon_0 > 0$. Assume the parameters $\alpha, k, \delta, \gamma$ underlying H_d satisfy the hypotheses of Lemma 5.10 and Lemma 4.2. Let θ be the value in Lemma 6.1, let t be as in Lemma 5.10 and set $r = \lceil 3 \log\log n \rceil$.*

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

2. *(Sandwiching.) For $U \sim U_n$ independent of $\vec{X}$,*
$$\mathop{\mathbb{E}}_{\vec{X},U}\left[u_{\phi,\vec{X}}(U) - \ell_{\phi,\vec{X}}(U)\right] \le 3s\varepsilon_0.$$

3. *(Simplicity.) Let $\Delta_0 = 40^4 \log^8(2n/\varepsilon_0)$. Then,*
$$\Pr_{\vec{X}}\left[\Delta\left(\ell_{\phi,\vec{X}}\right) \le \Delta_0 \text{ and } \Delta\left(u_{\phi,\vec{X}}\right) \le \Delta_0\right] \ge 1 - 2r\varepsilon_0.$$

Toward proving Lemma 6.2, fix a depth-$(d+1)$ read-once NAND formula ϕ, define $X_0 = \star^n$, and define $\ell_{\vec{X}}^{(0)} = u_{\vec{X}}^{(0)} = \phi$. Then, for $i < r$, inductively define
$$\ell_{\vec{X}}^{(i+1)} = \ell_{(\ell_{\vec{X}}^{(i)}|_{X_i})}.$$
That is, $\ell_{\vec{X}}^{(i+1)}$ is the lower sandwiching formula when Lemma 6.1 is applied to $\ell_{\vec{X}}^{(i)}|_{X_i}$. Similarly, define
$$u_{\vec{X}}^{(i+1)} = u_{(u_{\vec{X}}^{(i)}|_{X_i})},$$
i.e., $u_{\vec{X}}^{(i+1)}$ is the upper sandwiching formula when Lemma 6.1 is applied to $u_{\vec{X}}^{(i)}|_{X_i}$. Finally, define
$$\ell_{\phi,\vec{X}} = \ell_{\vec{X}}^{(r)}|_{X_r}$$
$$u_{\phi,\vec{X}} = u_{\vec{X}}^{(r)}|_{X_r}.$$

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

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

A completely analogous argument works for the upper bound as well. □

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

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

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

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

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 $i \ge 1$, the formula $\ell^{(i)}_{\vec{X}}$ and the formula $u^{(i)}_{\vec{X}}$ both have the property that every gate ψ satisfies $\mathbb{E}[\neg\psi] \ge \theta$, where

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

Furthermore, as the restrictions are independent, X_i is independent of $\left(\ell^{(i)}_{\vec{X}}, u^{(i)}_{\vec{X}}\right)$. Therefore, by Lemma 5.10,

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

and

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

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

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

and

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

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 $n^{(3/4)^i} \leq 10^4 \log^8(1/\theta)$. Then the bound becomes

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

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

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

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, $n^{(3/4)^r} \leq 2$, and by the definition of θ,

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

7 Fooling Formulas when Δ is Small

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

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

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 ([MRT18]). *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 constant-depth read-once 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 $\Delta(\phi) \leq \Delta_0$ 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 $\phi = \mathrm{NAND}(\varphi_1, \ldots, \varphi_m)$. 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 formulas over the $\{\wedge, \vee, \neg\}$ basis.

Lemma 8.1. *For any layered read-once 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

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

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,

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

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

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 5.4. 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 5.10. Let k, δ, γ be the values required by Lemma 4.2.

Correctness. Let ϕ be a depth-$(d+1)$ read-once formula. We can straightforwardly make ϕ a *layered* read-once 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 $X_1, \dots, X_r \sim H_d^{\circ t}$. 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,

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

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:

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

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

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

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 $d \geq 2$ 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 $s \leq (C_2 d \log\log(n/\varepsilon))^2$. 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

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

  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 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.

[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.

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

[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.

[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.

[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.

[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.

[MRT18] Raghu Meka, Omer Reingold, and Avishay Tal. Pseudorandom generators for width-3 branching programs. *arXiv preprint arXiv:1806.04256*, 2018.

[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 Bazzis theorem. *ACM Transactions on Computation Theory (TOCT)*, 1(1):3, 2009.

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

[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.

A Proof of Lemma 6.1

In this section, we give the proof of Lemma 6.1. We emphasize that this argument was already given by Chen, Steinke, and Vadhan [CSV15]; we are only reproducing it here to verify the exact parameters of Lemma 6.1.

Proof of Lemma 6.1. We proceed by induction on size(ϕ), i.e., the number of NAND gates, to prove the lemma with the modified bound $\mathbb{E}[u_\phi - \ell_\phi] \le n\sqrt{\theta} + \text{size}(\phi)\theta$. In the base case 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 $\phi = \text{NAND}(\phi_1, \dots, \phi_m)$. 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 $\ell_{\phi_i} \le \phi_i \le u_{\phi_i}$ with the following properties.

- $\mathbb{E}[u_{\phi_i} - \ell_{\phi_i}] \le n_i\sqrt{\theta} + \text{size}(\phi_i)\theta$.
- 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 $\mathbb{E}[\psi] \ge \theta$ and $\mathbb{E}[\neg\psi] \ge \theta$.

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

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

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

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

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 $n\sqrt{\theta} + \text{size}(\phi)\theta$. 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

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

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, $\mathbb{E}[u_\phi - \widetilde{\ell}_\phi] \leq n\sqrt{\theta} + \text{size}(\phi)\theta$. 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 $\ell_\phi = \text{NAND}(u_{\phi_i})$. 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

$$\ell_\phi = \text{NAND}(u_{\phi_1}, \dots, u_{\phi_j}).$$

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 $\text{size}(\phi) \leq n$ and $n\sqrt{\theta} + n\theta < \varepsilon$. □

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il