Fooling Constant-Depth Threshold Circuits (Extended Abstract)

Pooya Hatami
CSE Department
The Ohio State University
Columbus, OH, USA
pooyahat@gmail.com

William M. Hoza
Simons Institute
UC Berkeley
Berkeley, CA, USA
williamhoza@berkeley.edu

Avishay Tal
EECS Department
UC Berkeley
Berkeley, CA, USA
atal@berkeley.edu

Roei Tell
CSAIL
MIT
Cambridge, MA
roeitell@gmail.com

***Abstract*—We present new constructions of pseudorandom generators (PRGs) for two of the most widely studied non-uniform circuit classes in complexity theory. Our main result is a construction of the *first non-trivial PRG for linear threshold (*LTF*) circuits* of arbitrary constant depth and super-linear size. This PRG fools circuits with depth $d \in \mathbb{N}$ and $n^{1+\delta}$ wires, where $\delta = 2^{-O(d)}$, using seed length $O(n^{1-\delta})$ and with error $2^{-n^{\delta}}$. This tightly matches the best known lower bounds for this circuit class. As a consequence of our result, all the known hardness for LTF circuits has now effectively been translated into pseudorandomness. This brings the extensive effort in the last decade to construct PRGs and deterministic circuit-analysis algorithms for this class to the point where any subsequent improvement would yield breakthrough lower bounds.**

Our second contribution is a PRG for De Morgan formulas of size s whose seed length is $s^{1/3+o(1)} \cdot \mathrm{polylog}(1/\epsilon)$ for error ϵ. In particular, our PRG can fool formulas of sub-cubic size $s = n^{3-\Omega(1)}$ with an exponentially small error $\epsilon = \exp(-n^{\Omega(1)})$. This significantly improves the inverse-polynomial error of the previous state-of-the-art for such formulas by Impagliazzo, Meka, and Zuckerman (FOCS 2012, JACM 2019), and again tightly matches the best currently-known lower bounds for this class.

In both settings, a key ingredient in our constructions is a pseudorandom restriction procedure that has tiny failure probability, but simplifies the function to a non-natural "hybrid computational model" that combines several computational models. As part of our proofs we also construct "extremely low-error" PRGs for related circuit classes; for example, we construct a PRG for arbitrary functions of s LTFs that can handle even the extreme setting of parameters $s = n/\mathrm{polylog}(n)$ and $\epsilon = 2^{-n/\mathrm{polylog}(n)}$.

***Keywords*-pseudorandom generators; threshold circuits; De Morgan formulas**

I. Introduction

A pseudorandom generator (PRG) for a class $\mathcal{F}$ of functions $\{0,1\}^n \to \mathbb{R}$ is an efficient (deterministic) algorithm that maps a short random seed of length ℓ into a longer string of length n such that for every $f \in \mathcal{F}$,

$$\left| \mathop{\mathbb{E}}_{s \in \{0,1\}^{\ell}} [f(G(s)] - \mathop{\mathbb{E}}_{x \in \{0,1\}^{n}} [f(x)] \right| \leq \epsilon \,,$$

where ϵ is called the error of the PRG.

In this work we present new constructions of PRGs for two of the most widely studied non-uniform circuit classes in complexity theory. Our main result is a construction of the *first non-trivial PRG for linear threshold (*LTF*) circuits* of arbitrary constant depth and super-linear size. Prior to this work no non-trivial PRGs or deterministic satisfiability algorithms were known for LTF circuits of depth $d \geq 3$, and our result builds on considerable efforts dedicated to this challenge in the last decade. Moreover, our PRG is not only the first non-trivial one, but in fact already *tightly matches the best known lower bounds for* LTF *circuits* in terms of size and of error. Our second result is a PRG for De Morgan formulas of sub-cubic size that has an exponentially small error, where this error significantly improves on the previous state-of-the-art by [1]. In this setting too, the parameters of our PRG tightly match the best known lower bounds and correlation bounds for De Morgan formulas. Thus, in both settings, essentially any improvement in dependency of our PRGs on the circuit size or on the target error would improve the best known lower bounds for the corresponding circuit class.

A common initial technical challenge that underlies both of our PRGs is that of constructing *pseudorandom restriction procedures* that "simplify" the circuit with an *exponentially small failure probability.* The obstacle here is that the natural (and well-known) definitions of simplification do not yield such small failure probability, even if the restrictions were completely random. To overcome this obstacle, following [2], [3], [4], [5], we explore hybrid computational models that, despite being less natural, satisfy the following two competing properties: (1) They are strong enough so that the circuits simplify to those hybrid models except for an exponentially small probability; and (2) They are weak enough that we can fool them using a PRG with a suitable seed length. Our proofs hinge on a careful balance of this trade-off, as well as on PRG constructions for the corresponding hybrid model, both of which significantly improve on known technical results.

A. A PRG for super-linear size LTF circuits

Recall that a linear threshold function (LTF) is a Boolean function of the form $\Phi(x) = 1 \iff \sum_i w_i \cdot x_i > \theta$, where $w \in \mathbb{R}^n$ and $\theta \in \mathbb{R}$. The class of constant-depth linear-threshold circuits (LTF circuits) consists of circuits of constant depth whose gates can compute arbitrary LTFs.

2575-8454/21/$31.00 ©2021 IEEE
DOI 10.1109/FOCS52979.2021.00019

This circuit class has been studied since the '80s, both since it serves as a natural simple model of neural networks, and since it is a natural extension of circuit classes for which strong lower bounds have already been proved, such as AC^0 and $\mathsf{AC}^0[p]$.

While the common belief is that the class TC^0 of polynomial-sized constant-depth $\mathtt{LTF}$ circuits[1] is strictly weaker than the class NC^1 (of polynomial-sized De Morgan formulas), at the moment we do not even know of a function in $\mathsf{EXP}^{\mathsf{NP}}$ that is hard for TC^0. In fact, we do not even know lower bounds for $\mathtt{LTF}$ circuits of size (say) $n^{1.1}$: The best currently-known lower bounds against explicit functions were proved more than 25 years ago by Impagliazzo, Paturi, and Saks [8], who showed that the parity function is hard for $\mathtt{LTF}$ circuits of depth d with $n^{1+c^{-d}}$ wires, for some constant $c > 1$.[2] Despite the fact that no lower bounds for larger $\mathtt{LTF}$ circuits are known, *average-case* lower-bounds for circuits of the same size (up to the constant c) against functions in P were proved several years ago by Chen, Santhanam, and Srinivasan [3]. Also, for the special case of $d = 2$, Kane and Williams [13] proved that Andreev's function (which is in P) is hard for circuits with $n^{2.49}$ wires.

In the last decade, a line of works pioneered by Williams (see, e.g. [14], [15], [16], [17]) showed that lower bounds for a circuit class can be proved by constructing non-trivial deterministic circuit-analysis algorithms for circuits from this class; that is, by constructing algorithms for satisfiability or for $\mathtt{CAPP}$[3] that are faster than the trivial brute-force algorithm. Following Williams' [18] breakthrough lower bounds for ACC^0 circuits that relied on this approach, the natural subsequent major challenge in complexity theory is to try and finally prove better lower bounds for $\mathtt{LTF}$ circuits by constructing circuit-analysis algorithms for such circuits – see, e.g., the first open problem in [19], and also see [14], [20], [16], [21]. However, a major shortcoming is that so far we have not even been able to construct circuit-analysis algorithms that imply the *existing* lower bounds for $\mathtt{LTF}$ circuits from 1993, let alone new lower bounds; in other words, so far we have not even been able to "translate the known hardness into randomness".

Accordingly, in the past decade an extensive research effort has been devoted to this challenge, resulting in dozens of exciting works. For a single $\mathtt{LTF}$ (i.e., a single "gate" in the circuit), a long line-of-works culminated in a PRG with near-optimal seed length by Gopalan, Kane, and Meka (see [22], following [23], [24], [25], [26], [27], [28], [29], [30]). Various works constructed PRGs for "simple functions" of $\mathtt{LTF}$s, for example for $\mathtt{AND} \circ \mathtt{LTF}$ (aka polytopes, see [31], [25], [32], [33], [34], [35], [36]). For $\mathtt{LTF}$ circuits of *depth two* and subquadratic size, a PRG with seed length $n^{1-\Omega(1)}$ was constructed by Servedio and Tan [4]; and a satisfiability algorithm with running time $2^{n-n^{\Omega(1)}}$ was constructed by Alman, Chan, and Williams [37] (following [38], [39], [40]; this algorithm also works for the larger class $\mathsf{AC}^0[m] \circ \mathtt{LTF}_{n^{2-\Omega(1)}} \circ \mathtt{LTF}$, see Section III). However, despite these efforts, for circuits of *arbitrary depth* $d > 2$, prior to this work no non-trivial deterministic satisfiability or $\mathtt{CAPP}$ algorithm was known. The only known deterministic circuit-analysis algorithm for such circuits was an algorithm for the relaxed version of $\mathtt{CAPP}$ called *quantified derandomization* (see [41]), algorithms for which are not known to imply lower bounds. We defer further discussion of other relevant works to Section III.

Building on the rich ideas developed in the last decade, in this paper we construct the *first non-trivial PRG for* $\mathtt{LTF}$ *circuits* of arbitrary constant depth. Moreover, as we explain below, our PRG construction *tightly matches the best currently-known lower bounds* for such circuits – both the size lower bounds of [8] and the average-case lower bounds of [3]. Thus, our construction brings the extensive research effort described above to the point where essentially further improvement would yield new lower bounds for $\mathtt{LTF}$ circuits.

Theorem I.1 (PRG for super-linear $\mathtt{LTF}$ circuits)**.** *For any $d \in \mathbb{N}$ and $\delta \leq 200^{-d}$, there exists a polynomial-time computable ϵ-PRG for the class of $\mathtt{LTF}$ circuits of depth d with at most $n^{1+\delta}$ wires, whose seed length is $O(n^{1-\delta})$ and whose error is $\epsilon = 2^{-n^{\delta}}$.*

We comment that Theorem I.1 holds also for super-constant values of $d \in \mathbb{N}$ (for details, see the full version of the paper [42]). Parsing the parameters of our PRG, the seed length $O(n^{1-\delta})$ is "slightly non-trivial" (yielding a $\mathtt{CAPP}$ algorithm with running time $2^{O(n^{1-\delta})}$), yet essentially any improvement to this seed length would yield new size lower bounds for $\mathtt{LTF}$ circuits. Also, the error of our PRG is *exponentially small*, and again essentially any improvement to this error would imply new average-case lower bounds for $\mathtt{LTF}$ circuits (with respect to a natural polynomial-time-samplable distribution, as we explain in the full version [42]). It might seem surprising that the first non-trivial PRG already has such a small error, but this is not a coincidence: As explained above, a key technical challenge underlying our techniques is to reduce the error of certain auxiliary pseudorandom algorithms (i.e., of pseudorandom restriction procedures and of PRGs for a related class; we

[1] The class TC^0 is sometimes defined using unweighted majority gates and sometimes defined using $\mathtt{LTF}$ gates. Both definitions are equivalent up to polynomial overheads (see [6], [7]), but since we will be concerned with precise size bounds we will use a specific definition. Note that our PRG fools the stronger class.

[2] Here, by "explicit" we mean that these lower bounds are against functions in P. There are also incomparable lower bounds for *general circuits* (that in particular hold for $\mathtt{LTF}$ circuits) against functions that are "not very explicit", and in particular are not even known to be in NP (see, e.g., [9], [10], [11]). We also note that the precise value of the constant $c > 1$ in this expression turns out to be surprisingly important (see [12]).

[3] Recall that $\mathtt{CAPP}$ (the Circuit Acceptance Probability Problem) is the problem of distinguishing between circuits with acceptance probability at most $1/3$ and circuits with acceptance probability at least $2/3$.

elaborate on this in Section II-A).

As part of our proof of Theorem I.1 we also construct an "extremely-low-error" PRG for an *arbitrary function* of a bounded number of LTFs. In particular, our PRG fools any function of $s = n^{.99}$ LTFs with error $\epsilon = 2^{-n^{.99}}$ and seed length $n^{1-\Omega(1)}$; this setting of parameters is close to the maximal possible one (as there does not exist a non-trivial PRG for an arbitrary function of n variables, or with error 2^{-n}), and indeed we will use this PRG with such small error $\epsilon \approx 2^{-n^{.99}}$ in our proof of Theorem I.1. This significantly improves on the previous state-of-the-art, which could handle functions of $o(n^{2/5})$ LTFs and whose error is sub-exponential (see Section III for details). To present this result, for any $n, s \in \mathbb{N}$ denote by $\mathtt{ANY}_s \circ \mathtt{LTF}_n$ the class of functions $\{0,1\}^n \to \{0,1\}$ of the form $f(x) = g(\Phi_1(x), \ldots, \Phi_s(x))$, where the Φ_i's are LTFs and g is *arbitrary*. We prove that:

Theorem I.2 (low-error PRG for $\mathtt{ANY}_s \circ \mathtt{LTF}$)**.** *There exists an ϵ-PRG for $\mathtt{ANY}_s \circ \mathtt{LTF}_n$ that is computable in time* $\mathrm{poly}(n)$ *with seed length* $\widetilde{O}\Big(\sqrt{n \cdot (s + \log(1/\epsilon))}\Big)$.

One corollary of Theorem I.2 is a PRG with seed length $o(n)$ and error $\epsilon = 2^{-n/\mathrm{polylog}(n)}$ for the class of LTF circuits with *unbounded depth* and *at most* $\frac{n}{\mathrm{polylog}(n)}$ *gates* [42, Corollary 6.22]. The class of unbounded-depth LTF circuits has received less attention in recent years, compared to TC^0, and our PRG almost matches the $\Omega(n)$ lower bound that has been known for this class since the early '90s (see [43], [44], [45]).

A stronger efficiency requirement and a new lower bound: The PRG in Theorem I.1 in fact meets a stronger efficiency requirement than just being computable in polynomial time. Specifically, we show that the PRG can also be made *strongly explicit* (some sources use the term "local"): Given a seed s and an index $i \in [n]$, we can compute the i^{th} output-bit of the PRG $G(s)_i$ in time $O(|s|) = O(n^{1-\delta})$.

The existence of PRGs meeting such a strong efficiency requirement implies that the fooled circuit class cannot solve the Minimum Circuit Size Problem (MCSP) [46]. Thus, our results imply the first unconditional lower bound for solving MCSP by LTF circuits of super-linear size. (For context, recall that MCSP is widely believed to be hard even for P/poly, see [46], [47].) Moreover, our construction implies that such circuits cannot even solve the relaxed problem $\mathsf{gapMCSP}[s_1, s_2]$: In this promise problem we are given a truth-table $f \in \{0,1\}^{2^\ell}$ and need to decide whether the circuit complexity of f is at most $s_1(\ell)$ or at least $s_2(\ell)$.

Corollary I.3 (MCSP lower bound for LTF circuits of super-linear size; see [42, Theorem 6.26] for a more general statement)**.** *For any constant $d \in \mathbb{N}$ it holds that* $\mathsf{gapMCSP}\left[2^{(1-400^{-d})\cdot\ell}, 2^{\ell-1}/\ell\right]$ *cannot be decided by* LTF *circuits of depth d with $n^{1+400^{-d}}$ wires.*

The combination of Corollary I.3 and of recent "hardness magnification" results reveals a sharp threshold phenomenon for solving gapMCSP by LTF circuits of super-linear size. Specifically, improving the unconditional lower bound in Corollary I.3 to hold against *slightly* larger circuits would imply dramatic lower bounds for *all of* TC^0. This follows from the results of Chen, Jin, and Williams [48], which imply that for some constant $c > 1$, if for all $\beta > 0$ it holds that $\mathsf{gapMCSP}\left[2^{\beta\cdot\ell}, 2^{\ell-1}/\ell\right]$ cannot be decided by LTF circuits of depth $d' = 2d$ with $n^{1+c^{-d'}}$ wires, then NP is not contained in $\mathsf{TC}^0_d[n^k]$ for any fixed $k \in \mathbb{N}$. [4]

This sharp threshold phenomenon adds to several very recent results that demonstrated such a phenomenon for solving other problems by LTF circuits of super-linear size [12] (specifically, for solving certain NC^1 problems and for solving the problem of quantified derandomization), and for solving MCSP (or the closely related problem MKtP) by other circuit classes, including AC^0 circuits, $\mathsf{AC}^0[\oplus]$ circuits, and polynomial-sized formulas (see [49], [50], [48], [51], [52], [53]).

B. A low-error PRG for De Morgan formulas

Our second main result is a PRG for the class of De Morgan formulas, which consists of formulas of fan-in 2 over the De Morgan basis (i.e., with AND, OR, and NOT gates). This class has been widely studied since the early '60s, with a focus on the sub-class NC^1 of polynomial-sized formulas, which is a non-uniform analogue of computation in parallel logarithmic time.

A common conjecture is that NC^1 cannot compute all functions in P. However, at the moment, the best lower bounds that we know for NC^1 against explicit functions hold for De Morgan formulas of size $\frac{n^3}{\mathrm{polylog}(n)}$; these were proved by Håstad [54] (following [55], [56], [57], [58], [59]), with subsequent log-factor improvements [60], [61]. These bounds were extended to average-case lower bounds by Komargodski, Raz, and Tal [62] and Bogdanov [63] (following [64], [65]; see also [66], [61]), who showed that for any parameter $r \leq n$, De Morgan formulas of size $\frac{n^3}{r^2\cdot\mathrm{polylog}(n)}$ cannot compute a corresponding function in P with success probability more than $1/2 + 2^{-r}$; in particular, for $r = n^\delta$, this gives an average-case lower bound of $1/2 + 2^{-n^\delta}$ for De Morgan formulas of size $n^{3-2\delta-o(1)}$.

Almost a decade ago, Impagliazzo, Meka, and Zuckerman [1] were able to essentially match the known *formula size* lower bounds with a polynomial-time computable PRG, which has seed length $s^{1/3+o(1)}$, fooling De Morgan formulas of size s. While their PRG matches the known size lower bounds, it unfortunately supports only inverse-polynomial

[4] In fact, their result is even stronger, and only requires a lower bound against the non-gap version of MCSP for circuit-size $2^{\beta\cdot\ell}$ (see [48, Theorem 1.1, Item 7]. Thus, intuitively, the difference between the unconditional result in Corollary I.3 and a result that would imply lower bounds for all of TC^0 is even smaller.

error and not exponentially small error,[5] and therefore does not match the known *average-case* lower bounds, which assert at most an exponentially small advantage. Later on, Kabanets, Koroth, Lu, Myrisiotis, and Oliveira [36, Theorem 2] constructed a PRG for De Morgan formulas whose leaves are labeled by functions with low communication complexity. Their PRG fools a more general class, and its seed length has logarithmic dependency on the error parameter, but unfortunately the seed length is proportional to $\sqrt{s}$, and therefore this PRG is non-trivial only when the formulas are of quadratic size rather than of cubic size.

In this work we construct a PRG that nearly matches the known lower bounds for De Morgan formulas both in terms of formula size and in terms of the average-case hardness (i.e., in terms of the error probability of the PRG). In more detail:

Theorem I.4 (low-error PRG for De Morgan formulas)**.** *There exists a polynomial-time computable ϵ-PRG for De Morgan formulas of size s on n variables with seed length*

$$\left(s^{1/3} \cdot \log^{2/3}(1/\epsilon) + \log^2(1/\epsilon)\right) \cdot 2^{O(\sqrt{\log s})} \cdot \mathrm{polylog}(n) = s^{1/3+o(1)} \cdot \mathrm{polylog}(n/\epsilon) \,.$$

As one particular setting of the parameters, our PRG yields a function in NP that cannot be computed by De Morgan formulas of size $n^{3-2\delta-o(1)}$ with success probability more than $1/2 + 2^{n^{-\delta}}$ over a natural polynomial-time-samplable distribution, for any $\delta > 0$ (see [42, Proposition 4.11]). This essentially matches the best known average-case lower bounds for De Morgan formulas by [62], [63], which were mentioned above.

II. High-level proof overviews

We now present high-level overviews of our proofs. First, in Section II-A, we will describe the common high-level technical challenge underlying both constructions, and our general approach for handling this challenge. Then in Section II-B we describe our construction of a PRG for De Morgan formulas (i.e., Theorem I.4), which is considerably simpler than our PRG for LTF circuits and nevertheless showcases our approach. Finally, in Section II-C, we move on to the more involved construction of a PRG for LTF circuits (Theorem I.1).

A. The common high-level technical approach

Like most of the known unconditional PRGs for circuit classes, our constructions are based on *pseudorandom restrictions* that simplify every circuit in the class to a simpler circuit, with high probability. There are many known frameworks for obtaining PRGs from pseudorandom restrictions (see, e.g. [67], [1], [68], [69]), yet a common property is that the error of the PRG crucially depends on the *failure probability* of each restriction (i.e., the probability that the circuit does not simplify under restriction). In particular, when each restriction fails with probability p or more (where p is the fraction of variables kept alive by the restriction), we do not obtain any non-trivial PRG. (This is because these PRGs typically involve at least p^{-1} applications of restrictions.)

In classical analyses of restrictions (e.g., in [70], [54]), one aims to prove that every circuit simplifies to a circuit from the same class that is shallower or of significantly smaller size. The main problem for us is that such statements simply *do not hold with very high probability* for LTF circuits or for De Morgan formulas. For example, a size-n De Morgan formula might only depend on $O(\log n)$ variables. Under a random restriction, the formula remains completely intact with probability $p^{O(\log n)} > 2^{-O(\log(n)^2)}$. For LTF circuits the situation is even worse: Even a *single majority gate* fails to simplify with sufficiently high probability; we would like the gate to become constant under the restriction, or at least extremely biased, but the probability of that not happening is at least $\sqrt{p} \gg p$. This means that *we cannot hope to get any non-trivial PRG for* LTF *circuits using this approach*, and this has indeed been a main bottleneck prior to the current work.

Our way to bypass this obstacle in both settings, generalizing ideas from [2], [3], [4], [5], is to change the definition of what it means to "simplify". Instead of trying to claim that each circuit simplifies to a shallower or smaller circuit, as in classical results, we will claim that the restricted circuit can be computed by a hybrid computational model, which is an artificial combination of several models that is nevertheless "simpler" in some useful sense. Indeed, in both settings this relaxation allows us to reduce the failure probability of the restriction to be *exponentially small*. The trade-off, though, is that we will have to deal with restricted functions that are more complicated than just simpler circuits from the same class (i.e., they are computable by hybrid models). Our proofs will hinge on a careful balance of this trade-off. For example, improving on a previous result of [3], we will show that with probability $1 - \exp(-n^{\Omega(1)})$, restricted LTF circuits of super-linear size can be approximated by a decision tree of depth significantly less than $p \cdot n$ whose nodes query both variables and LTFs, and whose leaves are labeled by (small sets of) LTFs (see Proposition II.2 and the preceding explanation); indeed, the precise balance of parameters here is crucial for our PRG construction.

To be more specific, in each of the two settings we will need three new technical results to construct our PRG. First, we will show that a truly random restriction simplifies the circuit to a suitable hybrid model with probability $1 - \exp(-n^{\Omega(1)})$. Then, to use a restrictions-to-PRG framework, we will derandomize the latter result, showing

[5] More precisely, the result statements in [1] assert an error of $s^{-O(1)}$, but a careful examination of their analysis shows that an error of $1/s^{o(\log s)^{1/3}}$ is possible with similar seed length $s^{1/3+o(1)}$. Nonetheless, when the error is $1/s^{\omega(\log s)^{1/3}}$ the seed length becomes trivial.

that a suitable *pseudorandom restriction* also simplifies the circuit to the corresponding hybrid model with probability $1-\exp(-n^{\Omega(1)})$.[6] And lastly, we will have to fool the hybrid model in a way useful for the particular restrictions-to-PRG framework; for LTF circuits we construct a new PRG for the hybrid model (which will be a corollary of Theorem I.2), whereas for De Morgan formulas we will refine an extractor-based argument of [1] to work for the hybrid model.

We stress that our motivation for undertaking this approach is different in each of the two settings. For De Morgan formulas, we want to improve the error of the previous state-of-the-art PRG of [1]. However, for LTF circuits, as mentioned above, the failure probability of previously-known restrictions was a bottleneck toward obtaining *any* PRG whatsoever. Our motivation for reducing the failure probability is in order to construct the first non-trivial PRG for this class.

We comment that this general approach is also useful for fooling other circuit classes. For *branching programs* and for *formulas over an arbitrary basis*, it can provide PRGs with improved dependence on error compared to the previous state-of-the-art by Impagliazzo, Meka, and Zuckerman [1]. However, for these two classes, it turns out that a more elementary approach gives even better parameters, as we explain in the full version of the paper [42, Appendix B].

B. Low-error PRG for De Morgan formulas

For De Morgan formulas we will build on the PRG framework of Impagliazzo, Meka, and Zuckerman [1], which they used to construct the previous state-of-the-art PRG. Loosely speaking, their PRG framework combines $t \approx p^{-1}$ restrictions, and the PRG's error suffers a union-bound over the failure probability of these t restrictions. It is well-known that a random restriction shrinks every size-s De Morgan formula to a formula of expected size $O(p^2 \cdot s)$ (see [54], [60]), and in [1] they showed a concentration bound for this result that also holds for a pseudorandom restriction: For $p \geq 1/\sqrt{s}$, their restriction shrinks every size-s De Morgan formula to a formula of size $p^{2-o(1)} \cdot s$ with probability $1 - s^{-O(1)}$ (see [1, Lemma 4.8]).

As mentioned in Section II-A, it is impossible to improve the failure probability in their result to be smaller than $p^{O(\log(n))} > 2^{-\log(n)^2}$, since a De Morgan formula that is sensitive only to $O(\log(n))$ input variables does not simplify at all with such probability. Nevertheless, in this counterexample, a small number of variables are the ones responsible for the function's failure to simplify: In fact, if we were allowed to make a small number of "queries" to variables, the function would become trivial.

We show that *in general*, querying only a small number of variables helps us avoid almost all possible failure scenarios for the restriction: For any De Morgan formula of size s, with probability $1-\epsilon$ over a random restriction, the restricted formula can be ϵ-approximated by a decision tree of depth $s^{o(1)} \cdot \mathrm{polylog}(1/\epsilon)$ whose leaves are labeled by formulas of size $p^{2-o(1)} \cdot s$. Moreover, we show that this happens also for a suitable pseudorandom restriction:

Proposition II.1 (low-error pseudorandom restrictions for De Morgan formulas, informal)**.** *For any $n, s \in \mathbb{N}$, $p \in (1/n, 1/2)$ and $\epsilon > 0$, there exists a distribution over restrictions $\rho \in \{0,1,\star\}^n$ keeping each variable alive with marginal probability $p' \geq p$ that is samplable in time $\mathrm{poly}(n,s)$ with $s^{o(1)} \cdot \mathrm{polylog}(n/\epsilon)$ random bits and satisfies the following. For every size-s De Morgan formula f, with probability at least $1-\epsilon$ the formula $f\restriction_\rho$ can be ϵ-approximated by a decision tree of depth $s^{o(1)} \cdot \mathrm{polylog}(sn/\epsilon)$ with formulas of size $p^{2-o(1)} \cdot s$ at its leaves.*[7]

Let us first describe the main idea in the proof of Proposition II.1. Recall that a De Morgan formula is called read-k if each variable appears at most k times among the leaves. In [1] they first showed that *read-k formulas* shrink with extremely high probability; specifically, for $k = \frac{p^{O(1)}}{\log(s/\epsilon)} \cdot s$, they showed that a pseudorandom restriction shrinks any read-k formula from size s to size $O(p^2 \cdot s)$, with probability $1-\epsilon$. This can indeed yield an exponentially small error with seed length smaller than n, and the main part in their analysis that increases the error to $1/\mathrm{poly}(s)$ is a subtle reduction from the case of general De Morgan formulas to the case of read-k De Morgan formulas. (Similarly, the analyses of [65], [62], [71] also had to handle the "heavy" variables in a non-trivial manner.)

Our key observation here is simple: Using a DT, we can just *query all the "heavy" variables*, i.e., variables that appear more than k times, thereby reducing the case of a general De Morgan formula to the case of a DT with read-k De Morgan formulas at its leaves. Since there are at most s/k heavy variables, the depth of our DT will be at most $s/k = \mathrm{poly}(p^{-1}) \cdot \log(s/\epsilon)$. While this does not yet achieve the parameters stated in Proposition II.1, we follow [1] in composing less than $\log(1/p)$ restrictions that each keep a $q = s^{-o(1)}$ fraction of live variables such that their composition keeps a p fraction of live variables; the depth of our DT is thus less than $\log(1/p) \cdot \mathrm{poly}(q^{-1}) \cdot \log(s/\epsilon) < s^{o(1)} \cdot \mathrm{polylog}(sn/\epsilon)$, as stated in Proposition II.1.

The trade-off, however, is that since we simplify a De Morgan formula to a hybrid model rather than to a smaller De Morgan formula, a naive application of the PRG framework of [1] would yield a trivial seed length: This is because the seed length in their analysis is proportional to the description length of the restricted function, whereas our

[6] We note in advance that our technical result statements typically already assert the result for a pseudorandom restriction (which is stronger than the corresponding result for a random restriction).

[7] To use this result in our PRG construction we actually need a stronger notion of approximation. In our technical result we show that the formula is approximated with "zero-error" by the hybrid model, but for simplicity we ignore this in the high-level overview.

hybrid model requires a very large description (exponential in its depth). To overcome this we modify their analysis such that it can handle our hybrid model. Specifically, we show that if the restricted function can be computed by a DT with m leaves, each labeled with a function of description length s_0, then we can replace an additive term of $\widetilde{O}(m \cdot s_0)$ in the seed length (which is too much for us) with an additive term of $\widetilde{O}(s_0 + \log(m))$, at the (mild) cost of multiplying the final seed length by $\log(m/\epsilon)$. See the full version of the paper [42] for details.

C. PRG for super-linear LTF circuits

We now describe the proof of Theorem I.1. For simplicity, in the high-level overview we think of $d \in \mathbb{N}$ as a constant, and fix $\delta = 2^{-O(d)}$, where the O hides a universal constant. We want to construct a PRG for LTF circuits of depth d with at most $n^{1+\delta}$ wires, which has seed length $n^{1-\delta}$ and error $2^{-n^{\delta}}$. As part of this proof we will also describe the proof of Theorem I.2 (our PRG for $\mathtt{ANY}_s \circ \mathtt{LTF}_n$), and a self-contained description of the latter appears in Section II-C3.

1) Overview: Basic ideas and main challenges: For this setting we will use the classical restrictions-to-PRG framework of Ajtai and Wigderson [67]. The first component needed to instantiate this framework is a pseudorandom restriction, or more specifically a pseudorandom way to choose $\approx p \cdot n$ variables such that for every LTF circuit with depth d and $n^{1+\delta}$ wires, when fixing the rest of the variables *uniformly*, with high probability the circuit simplifies to some class $\mathsf{C}_{\mathtt{simple}}$. The second component that we need is a PRG for the class $\mathsf{C}_{\mathtt{simple}}$.

Random restrictions for LTF circuits of depth d with $n^{1+\delta}$ wires were previously studied in [72], [3]. In the most relevant result to our setting, Chen, Santhanam, and Srinivasan [3, Lemma 39] proved that a random restriction simplifies any such circuit to a corresponding hybrid model with *exponentially small* failure probability (jumping ahead, the hybrid model that we will use will be a refinement of their hybrid model). Moreover, even a *pseudorandom* restriction procedure for such circuits is already known (see [41]). The foregoing procedures (as well as all other procedures that we will mention below) use the parameter value $p = n^{-\alpha}$, where α is a small constant. However, these restriction procedures do not suffice in order to obtain a PRG via the [67] framework. Concretely, we are faced with three main challenges:

1) **Stronger simplification of the restricted function.** The first challenge is that in previous analyses *the hybrid model to which the restricted* LTF *circuit simplifies is not "simple enough"* to be useful in known restrictions-to-PRG frameworks. Specifically, to get a PRG we will need to "fool" the restricted circuit using significantly less randomness than the remaining $p \cdot n$ bits. However, in [3], [41] the hybrid model involves a DT of depth $(1 - o(1)) \cdot (p \cdot n)$, which requires seed length essentially $p \cdot n$ to "fool".[8] We need to show that random restrictions (and, later on, pseudorandom ones) simplify any LTF circuit to a "sufficiently simple" hybrid model, for which we can (potentially) construct an unconditional PRG with seed length $o(p \cdot n)$.
2) **Low-error derandomization.** The second challenge is that *the error probability of the known pseudorandom restriction procedure is too large* to be useful in the known restrictions-to-PRG frameworks. As mentioned in Section II-A, the [67] framework involves a union-bound over p^{-1} restrictions, and therefore the error of each restriction has to be at most p. However, the analysis of pseudorandom restrictions in [41] only bounds the error by $p^{1/5}$, using a naive concentration bound (i.e., Markov's inequality); whereas the analysis of [3] for truly uniform restrictions relies on a read-k Chernoff bound (i.e., on [73]), which is not known to hold for a suitable pseudorandom distribution.
3) **Constructing a PRG for the hybrid model.** Lastly, after we show that suitable pseudorandom restrictions simplify any LTF circuit to a "sufficiently simple" hybrid model with sufficiently small failure probability, we need to *construct a PRG with seed length* $o(p \cdot n)$ and error smaller than p for the hybrid model. As we will explain in Section II-C3, previously-known PRG constructions do not seem to suffice for this purpose.

We now state our two key technical results underlying Theorem I.1, corresponding to the challenges above. First, we construct a pseudorandom restriction procedure with seed length approximately p^{-1} and failure probability $\epsilon = 2^{-n^{\delta}}$ that simplifies any LTF circuit of super-linear size to a sufficiently simple hybrid model. In more detail, the hybrid model that we consider is a DT whose gates query *both* LTF*s and variables*, with no more than $p^{\Omega(1)} \cdot (p \cdot n)$ variables and $O(n^{1/4})$ LTFs queried in each path, and whose leaves are labeled by LTFs. Indeed, the precise depth and number of queries of each type that this DT makes are of crucial importance to our results. (Our actual hybrid model is unfortunately slightly more complicated, labelling each leaf with a small set of LTFs rather than with a single LTF, since for our PRG we will need to show that any LTF circuit can be *sandwiched with error* ϵ between two functions that are each computable by such a hybrid model. For simplicity, we ignore this fact and the more complicated model in the high-level overview.)

Proposition II.2 (low-error pseudorandom restrictions for super-linear LTF circuits, informal)**.** *For any constant* $d \in \mathbb{N}$ *and* $\delta = \frac{1}{2} \cdot 50^{-d}$*, there is a distribution over subsets* $\mathbf{I} \subseteq [n]$

[8] In [41], the pseudorandom algorithm gets as input an LTF circuit and queries variables according to that specific circuit, but this argument can be easily converted to a "black-box" pseudorandom restriction algorithm that simplifies any circuit to a DT with parameters essentially as in [3].

of size $\lceil pn \rceil$, where $p = n^{-(1+\delta)/10}$, that can be sampled in time $\mathrm{poly}(n)$ with $n^{1/10+O(\delta)}$ random bits, such that the following holds. For any depth-d LTF circuit over n bits and with $n^{1+\delta}$ wires, when fixing uniform values for the variables in $\lceil n \rceil \setminus \mathbf{I}$, with probability at least $1 - 2^{-n^{\delta}}$ the restricted circuit can be $2^{-n^{\delta}}$-approximated by a decision tree in which each path queries at most $p^{\Omega(1)} \cdot (p \cdot n)$ variables and $O(n^{1/4})$ LTFs, and each leaf is labeled by an LTF.

Our second key technical result is a low-error PRG for the hybrid model from Proposition II.2, which has seed length $p^{\Omega(1)} \cdot (p \cdot n)$ (note that this is essentially the best possible seed length, given that the DT queries $p^{\Omega(1)} \cdot (p \cdot n)$ variables in each path). This low-error PRG will follow as a special case of the PRG that was stated in Theorem I.2.

Proposition II.3 (low-error PRG for the hybrid model, informal)**.** *Consider the class of functions over n' input bits that are computable by decision trees that in each path query at most D variables and M LTF functions, and whose leaves are labeled by LTFs. Then, there exists an ϵ-PRG for this class, computable in $\mathrm{poly}(n')$ time, with seed length $\widetilde{O}\Big(\sqrt{n' \cdot (D + M + \log(1/\epsilon))}\Big)$.*

In our application, given the restriction procedure in Proposition II.2, we will have $n' = \lceil pn \rceil$ and $D = p^{\Omega(1)} \cdot (p \cdot n)$ and $M = O(n^{1/4})$, and we will use the error parameter $\epsilon = 2^{-n^{\delta}}$. Therefore, the seed length of the PRG from Proposition II.3 will be dominated by $\widetilde{O}(\sqrt{(p \cdot n) \cdot D}) \leq p^{\Omega(1)} \cdot (p \cdot n)$.

In the following Sections II-C2 and II-C3 we will describe the main ideas behind the proofs of Propositions II.2 and II.3, respectively. We note that these two sections can be read independently of each other.

2) Low-error pseudorandom restrictions that "sufficiently simplify" the circuit: We now describe the proof of Proposition II.2, in high-level and while not specifying precise parameter values for simplicity. We will iteratively restrict the circuit for $d - 1$ iterations; in each iteration i we start with a DT whose leaves are labeled by LTF circuits of depth i, and our goal is to simplify it to a DT whose leaves are labeled by LTF circuits of depth $i - 1$. For simplicity, let us first ignore the parameters of the DT, and just focus on a single circuit.

A single iteration: We choose the variables to keep alive via a k-wise independent distribution, for $k \approx p^{-1} \cdot \log(1/\epsilon)$. Following [3], [41], we partition the graph between the gates at the bottom layer and the variables into three parts: The one induced by "heavy" variables, the one between "light" gates and "light" variables, and the remaining one between "heavy" gates and "light" gates (we intentionally avoid precise definitions in this high-level description). Our goal is to show that after the restriction, and given appropriate queries of variables and of LTFs by the DT, all light gates will have fan-in at most one, and all heavy gates will become extremely biased. In this case we will replace the heavy gates by the corresponding constant, and will thus be able to reduce the depth of the circuit by one (at a cost of a small approximation error).

1. Heavy variables. Analogously to the setting of De Morgan formulas, our DT first queries all the heavy variables. Recall that the circuit has only $n^{1+\delta}$ wires; we define heavy variables so that the DT would query at most $p^{\Omega(1)} \cdot (p \cdot n)$ such variables.

2. Light gates and light variables. The subgraph induced by light gates and light variables was handled in previous arguments using a simple graph-theoretic argument, which resulted in a DT that is too deep for our purposes (i.e., the previous DTs were of depth $(1 - o(1)) \cdot (p \cdot n)$ whereas we need depth $o(p \cdot n)$). We handle this subgraph using a more refined graph-theoretic argument. First, we carefully set the parameters (in all other parts of our proof) such that the expected number of variable-pairs in this subgraph that both feed into a common gate and that survive the restriction is $p^{\Omega(1)} \cdot (p \cdot n)$.

Now we prove a concentration bound, showing that with probability $1 - \epsilon$ under our choice of restrictions, indeed at most $p^{\Omega(1)} \cdot (p \cdot n)$ such variable-pairs survive the restriction. To prove this bound we rely on the fact that the subgraph between light gates and light variables has *small degree*: This allows us to partition the light gates into few large sets that read disjoint subsets of variables. Given this concentration bound, with probability $1 - \epsilon$, after the restriction our DT can query all the $p^{\Omega(1)} \cdot (p \cdot n)$ living variables participating in such pairs, hence reducing the fan-in of all gates in the subgraph to at most one (which allows us to merge these gates into the layer above them).

3. Heavy gates and light variables. Lastly, we are left with the subgraph between heavy gates and light variables, which is the most interesting part in the argument. The analysis of [3] for a truly random restriction handled this subgraph with an exponentially small failure probability; but this analysis relied on a read-k Chernoff bound [73], which we do not know how to derandomize in our particular setting using only $p \cdot n$ random bits. We use a k-wise independent choice of variables to keep alive, and rely on an analysis that refers to the particular structure of each LTF function (computed by a gate in the circuit) to show that with all but an exponentially small failure probability, we can simplify this subgraph after at most $p^{\Omega(1)} \cdot (p \cdot n)$ queries to variables and $p^{-O(1)}$ queries to LTFs. Details follow.

The idea underlying previous results is to rely on a "restriction lemma" for a single LTF, which shows that each gate in this subgraph becomes extremely biased with probability $1 - p^{\Omega(1)}$ when restricted. Thus, we expect the fan-in of each gate in this subgraph to decrease by a factor of about p (recall that gates are heavy), and that all but a

$p^{\Omega(1)}$ fraction of the gates will become extremely biased. When this happens, we can replace the extremely biased gates by constants, thereby reducing the number of wires in the subgraph by a $p \cdot p^{\Omega(1)}$ factor, and then we can query of all the $p^{1+\Omega(1)} \cdot n^{1+\delta} = p^{\Omega(1)} \cdot (p \cdot n)$ remaining variables in the subgraph using our DT (hence eliminating the subgraph completely). However, it is not clear how to show that the decrease of $p^{1+\Omega(1)}$ in the number of wires happens with high probability, rather than just in expectation.

Recall that our choice of values for fixed variables is *uniform*, but that our choice of which variables to keep alive is only k-wise independent. The key problem is the latter choice might restrict some gates in a manner such that we can no longer claim that a uniform choice of values makes these gates biased with probability $1 - p^{\Omega(1)}$. To overcome this problem, we prove that with all but exponentially small probability, after choosing the live variables we can use the DT to *query* $p^{\Omega(1)} \cdot (p \cdot n)$ *additional variables in a careful way, which takes into account the particular structure of each* LTF *gate*, such that after these queries, each LTF becomes biased with probability at least $1 - p^{\Omega(1)}$ over a uniform choice of values for the restricted variables. We stress that we are considering two different events and distributions here: We are interested in proving that with extremely high probability $1-\epsilon$, our pseudorandom choice of variables is "good" for each and every LTF gate (after querying additional variables); whereas the meaning of "good" here is that with moderately high probability $1 - p^{\Omega(1)}$ over random choice of values for fixed variables, the LTF gate becomes biased. Conditioned on any successful choice of live variables, we can *now* apply the read-k Chernoff bound to the uniform choice of values for fixed variables, and deduce that the fraction of unbiased gates is very close to $p^{\Omega(1)}$. One caveat is that during this process, our DT will also query a small number of LTFs, rather than only variables.

Subsequent iterations and approximation errors: The above procedure transforms a circuit C_d of depth d into a DT over LTFs and variables whose leaves are labeled by circuits of depth $d-1$ and that *approximates* C_d with very small error, where the approximation error comes from the fact that we replaced biased gates by constants.

Our goal now is to iteratively apply further restrictions, in order to further reduce the depth of the LTF circuits at the leaves of the DT, until we reach a DT whose leaves are labeled with LTF circuits of depth one (i.e., LTFs). For $i = d-1, \ldots, 1$, we reduce the model to a decision tree querying at most D_i variables and M_i gates, and most importantly, whose leaves are circuits of depth i. (We index iterations backwards as they correspond to the depth of the LTF circuits on the leaves.) Note that when applying a restriction with value p_i to a DT of depth D_i, in addition to claiming that the LTF circuits at the leaves of the DT become shallower, we also need to claim that the depth of the tree itself decreases to roughly $p_i \cdot D_i$ (to ensure that the final depth of the DT is sub-linear in the number of alive variables). We show that in each iteration both statements hold for $1-\epsilon$ of the leaves.

However, when composing restrictions in this manner we are faced with a subtle issue, which is the bottleneck in the proof that *necessitates having an exponentially small error in each restriction* (i.e., the argument would not follow through with larger error). Recall that each leaf contributes a small error to the global tree, where the source of error is that the new DT that labels this leafs only approximates the corresponding function. Also recall that when counting the global error, the underlying distribution refers to the errors each leaf makes on inputs that correspond to this leaf, under a uniform choice of input. The issue arises since the initial DT queries not only variables but also LTFs: Hence, a uniform choice of input does not induce a uniform choice of input *in each leaf*, since the set of inputs that reach any particular leaf are the ones who also satisfy the queries of the LTF gates along the path. In particular, this means that the weight of errors inside each leaf might be amplified.

The key to resolving this issue is to rely on the fact there are at most M_i LTFs in each path, and therefore we intuitively expect the distribution over inputs inside the leaf to be skewed by a multiplicative factor of at most 2^{M_i}. We indeed formalize this intuition, and to solve the issue, in each restriction i we ensure that the number of queried LTFs is at most $M_i = p_i^{-O(1)}$, and we make sure that *the error in the subsequent iteration*, ϵ_{i-1} will be much smaller than 2^{-M_i}. (This is done by choosing, in each subsequent iteration, a smaller value for p_{i-1}, i.e., $p_{i-1} \ll p_i$.) Hence, the global in the subsequent restriction will be at most $\epsilon_{i-1} \cdot 2^{M_i} \ll \epsilon_i$.

3) Low-error PRG for the "sufficiently simple" hybrid model: Our goal in this section is to prove Theorem I.2, i.e., to construct a PRG for the class $\texttt{ANY}_s \circ \texttt{LTF}_n$ of functions that can be computed as an arbitrary function of s LTFs, whose seed length is $\widetilde{O}\Big(\sqrt{n \cdot (s + \log(1/\epsilon))}\Big)$. Our main application of this result is Proposition II.3, which follows easily as a corollary (see [42, Section 2.3.3] for an explanation). We note in advance that the crucial thing for this corollary is that the PRG will be able to handle a tiny error of $\epsilon \approx 2^{-n^{.99}}$.

Until recently, the seed length of known PRGs, even for the special case of $\texttt{AND} \circ \texttt{LTF}$, was proportional to $\log(1/\epsilon)^2$, which is too much for us (see [31], [32], [33], [35]). However, very recently Kabanets, Koroth, Lu, Myrisiotis, and Oliveira [36] constructed a PRG that has a better dependency on the error, while simultaneously handling a larger class of composition functions. Specifically, when the composition function is a De Morgan formula of size s, their seed length is $\widetilde{O}\big(\sqrt{n} \cdot s^{1/4} \cdot \log(1/\epsilon)\big)$. While this is still not good enough for our application, their ideas will serve as our starting point.

Let $f(x) = h(g_1(x), \ldots, g_s(x))$ for g_i's that are LTFs and for some composition function h. Informally, the main

idea underlying [36] is to reduce the problem of ϵ-fooling f to the problem of δ-fooling communication protocols for functions of the form $\bar{g}(x) = \prod_{j\in[\Delta]} g_{i_j}(x)$, where $\Delta \in \mathbb{N}$ is not too large but the error δ is very small. To do so, in the analysis they first $(\epsilon/3)$-approximate h by a real polynomial p_h of bounded degree Δ, then replace each of the monomials $\bar{g}$ of the polynomial by a corresponding randomized communication protocol with error $\epsilon/3s$, and finally claim that our PRG "fools" each of the communication protocols with sufficiently low error $\delta \ll \epsilon/2^{\widetilde{O}(\Delta)}$ allowing for a union-bound over monomials. (See [36, Theorem 25] for details.)

Instantiating the approach above with the trivial degree-s polynomial representation of h and with efficient randomized communication protocols for functions of the form $\bar{g}$ and with suitable PRGs for these protocols, the resulting seed length is $\widetilde{O}\big(\sqrt{n\cdot s}\cdot\log(1/\epsilon)\big)$. Tracking the parameters carefully, the multiplicative term of $\log(1/\epsilon)$ comes from computing each $\bar{g}$ up to error $\epsilon/3s$.

Our main idea is to avoid the multiplicative overhead of $\log(1/\epsilon)$ by *making the polynomial p_h "robust to noise" at each coordinate*, which allows us to use *communication protocols with constant error rather than with error $\epsilon/3s$*. To do so we use a beautiful result of Sherstov [74]: For every polynomial p_h, he constructed a "robust" polynomial $\widetilde{p_h}$ of degree $d = O(\deg(p_h)+\log(1/\epsilon))$ such that for every input $x \in \{0,1\}^n$ and "noise" $\eta \in [-1/3,1/3]^n$ it holds that $\Big|p_h(x) - \widetilde{p_h}(x+\eta)\Big| < \epsilon$. In high-level, to ϵ-approximate $h(g_1,\ldots,g_s)$ by a low-degree polynomial of communication protocols, we will take a trivial representation of h by a degree-s polynomial p_h, convert p_h to the robust polynomial $\widetilde{p_h}$ guaranteed by [74], and instead of "feeding" $\widetilde{p_h}$ the functions $g_1,\ldots,g_s$, we will feed $\widetilde{p_h}$ the *expected values of the communication protocols for each g_i*, while relying on the fact that $\widetilde{p_h}$ is robust to the errors of the protocols.

In more detail, for each g_i denote by $\mathbf{g}_i$ a randomized communication protocol for g_i with error $1/3$. Then, for every $x \in \{0,1\}^n$ we have that

$$\Big|h(g_1(x),...,g_s(x)) - \widetilde{p_h}(\mathbb{E}[\mathbf{g}_1(x)],...,\mathbb{E}[\mathbf{g}_s(x)])\Big| < \epsilon/3\,,$$

where we relied on the fact that for each i it holds that $\mathbb{E}[\mathbf{g}_i(x)]$ is $(1/3)$-close to $g_i(x)$ and that $\widetilde{p_h}$ is $(\epsilon/3)$-robust to a noise of up to $1/3$ per coordinate. Our goal is to fool the function $\widetilde{f}(x) = \widetilde{p_h}(\mathbb{E}[\mathbf{g}_1(x)],...,\mathbb{E}[\mathbf{g}_s(x)])$, and we want to show that it suffices to use a PRG that δ-fools communication protocols for functions of the form $\bar{g} = \prod_i g_i$, where δ is sufficiently small. The final observation that allows us to do so is that any monomial of $\widetilde{f}$, which is of the form $\Pi_i\,\mathbb{E}[\mathbf{g}_i]$ can be thought of as the expected value of the natural randomized protocol that independently runs protocols for the g_i's and accepts if all of the protocols accepts. Since any PRG for communication protocols also fools the expected value of a randomized protocol, our PRG fools the monomials of $\widetilde{f}$ with low error. Assuming that the error is sufficiently small to allow for a union-bound over monomials (taking into account the weights of their coefficients), our PRG also fools $\widetilde{f}$ itself.

The argument above allows us to replace the $(\epsilon/3s)$-error of the communication protocols by error $\rho = 1/3$. We then instantiate communication protocols (for composition of d LTFs) and PRGs (for the communication protocols) as above, and obtain a PRG for $\mathtt{ANY}_s\circ\mathtt{LTF}$ with seed length $\widetilde{O}(\sqrt{n\cdot d}\cdot\log(1/\rho)) = \widetilde{O}(\sqrt{n\cdot(s+\log(1/\epsilon))})$.

III. Previous work on circuit-analysis algorithms for LTF circuits

As mentioned in Section I-A, a large number of previous works focused on circuit-analysis algorithms for LTF circuits, and we now survey the previously known algorithms that are deterministic. (We do not survey the many known *randomized* algorithms here; see, e.g., [3], [37], [75], [36].)

Single LTFs and simple compositions of LTFs: For a single LTF, a PRG with near-optimal seed length $\widetilde{O}(\log(n/\epsilon))$ was constructed by Gopalan, Kane, and Meka [22], following [23], [24], [25], [26], [27], [28], [29], [30]. Concurrently and subsequently, various PRGs were constructed for "simple compositions" of LTFs, and in particular for $\mathtt{AND}\circ\mathtt{LTF}$ (i.e., for polytopes, see [31], [25], [32], [33], [34], [35], [36]), for monotone functions of LTFs [31], and for small De Morgan formulas of LTFs [36].

The class $\mathtt{ANY}_{o(n)}\circ\mathtt{LTF}$: The problem of fooling $\mathtt{ANY}_s\circ\mathtt{LTF}$ with error ϵ reduces to fooling $\mathtt{AND}_s\circ\mathtt{LTF}$ with error $\epsilon/2^s$ [34, Footnote 1]. Combining this reduction with the PRG of [36, Theorem 30], one can obtain a PRG for $\mathtt{ANY}_s\circ\mathtt{LTF}$ with seed length $\widetilde{O}\big(\sqrt{n}\cdot(s^{5/4}+s^{1/4}\cdot\log(1/\epsilon))\big)$, which is non-trivial for $s \le n^{2/5}/\mathrm{polylog}(n)$ (note that this result is superseded by Theorem I.2). Chattopadhyay, De, and Servedio [34] (following [31]) constructed a deterministic algorithm that approximately counts the fraction of satisfying assignment for a given $\mathtt{ANY}_s\circ\mathtt{LTF}$ circuit, up to error ϵ, in time $\mathrm{poly}(n)\cdot 2^{\mathrm{poly}(s,1/\epsilon)}$. Note that their running time has an optimal dependency on n, but becomes trivial when $s \ge n^{\Omega(1)}$ or $\epsilon \le n^{-\Omega(1)}$. In comparison, the PRG from Theorem I.2 always has seed length at least $\sqrt{n}$, which is sub-optimal in the parameter n, but its seed length remains $o(n)$ even for $s = n/\mathrm{polylog}(n)$ and $\epsilon = 2^{-n/\mathrm{polylog}(n)}$.

Constant-depth LTF circuits: For circuits of depth two (i.e., $\mathtt{LTF}\circ\mathtt{LTF}$ circuits), Servedio and Tan [4] constructed an $(n^{-O(1)})$-PRG with seed length $n^{1-\Omega(1)}$ that works when the number of wires is subquadratic. In an incomparable result, Alman, Chan, and Williams [37] (following [38], [39], [40]) constructed a satisfiability algorithm that runs in time $2^{n-n^{\Omega(1)}}$ for the larger class of $\mathsf{AC}^0[m]\circ\mathtt{LTF}\circ\mathtt{LTF}$ circuits of *subexponential size* that have a subquadratic number of LTF gates at their bottom layer.

For LTF circuits of depth $d > 2$, prior to the current work the known PRGs and satisfiability algorithms only handled circuits with at most $n^{.49}$ gates. However, these works

extended to the more general model of AC^0 circuits that are augmented by a bounded number LTF gates: Specifically, Servedio and Tan [76] (following [77]) constructed a PRG for AC^0 circuits of size S with at most $2^{\alpha \cdot \sqrt{\log S}}$ LTF gates (for a universal constant $\alpha > 0$) whose seed length is $2^{O(\sqrt{\log S})} + \mathrm{polylog}(1/\epsilon)$; and Lovett and Srinivasan [78] constructed an incomparable PRG for AC^0 circuits of polynomial size with at most $n^{\cdot 49}$ LTF gates whose seed length is n^{δ} (for an arbitrarily small $\delta > 0$) and whose error is $2^{-n^{\cdot 24}}$. See [79] for another result in this spirit.

The only previously-known algorithm for LTF circuits of depth $d > 2$ and super-linear size was an algorithm for quantified derandomization (i.e., for a relaxed circuit-analysis task) that runs in time $n^{\mathrm{polyloglog}(n)}$ and works when the circuit has $n^{1+2^{-O(d)}}$ wires and evaluates to the same output on all but $2^{n^{1-2^{-O(d)}}}$ of its inputs [41].

ACKNOWLEDGMENTS

We are grateful to Lijie Chen, Dean Doron, Li-Yang Tan, and David Zuckerman for very helpful discussions. In particular, we thank Lijie for suggesting that we consider implications of our construction to MCSP lower bounds. P.H. is supported by NSF grant CCF-1947546. When this research was conducted, W.M.H. was a graduate student at the University of Texas at Austin, supported by the NSF GRFP under Grant DGE-1610403 and by a Harrington Fellowship from UT Austin. Part of this work was done while R.T. was supported by funding from the European Research Council (ERC, grant agreement No. 819702), and while he was a postdoctoral fellow at DIMACS and supported by the NSF grant number CCF-1445755.

REFERENCES

[1] R. Impagliazzo, R. Meka, and D. Zuckerman, "Pseudorandomness from shrinkage," *J. ACM*, vol. 66, no. 2, pp. 1–16 (Art. 11), 2019.

[2] J. Håstad, "On the correlation of parity and small-depth circuits," *SIAM J. Comput.*, vol. 43, no. 5, pp. 1699–1708, 2014.

[3] R. Chen, R. Santhanam, and S. Srinivasan, "Average-case lower bounds and satisfiability algorithms for small threshold circuits," *Theory Comput.*, vol. 14, pp. 1–55 (Art. 9), 2018.

[4] R. Servedio and L.-Y. Tan, "Learning and fooling depth-two threshold circuits," 2017, unpublished.

[5] A. Tal, "Tight bounds on the Fourier spectrum of AC0," in *Proc. 32nd Annual IEEE Conference on Computational Complexity (CCC)*, 2017, pp. 15:1–15:31.

[6] M. Goldmann, J. Håstad, and A. Razborov, "Majority gates vs. general weighted threshold gates," in *Proc. 7th Annual Structure in Complexity Theory Conference*, 1992, pp. 2–13.

[7] M. Goldmann and M. Karpinski, "Simulating threshold circuits by majority circuits," *SIAM J. Comput.*, vol. 27, no. 1, pp. 230–246, 1998.

[8] R. Impagliazzo, R. Paturi, and M. E. Saks, "Size-depth tradeoffs for threshold circuits," in *Proc. 25th Annual ACM Symposium on Theory of Computing (STOC)*, 1993, pp. 541–550.

[9] R. Kannan, "Circuit-size lower bounds and non-reducibility to sparse sets," *Information and Control*, vol. 55, no. 1-3, pp. 40–56, 1982.

[10] H. Buhrman, L. Fortnow, and T. Thierauf, "Nonrelativizing separations," in *Proc. 13th Annual IEEE Conference on Computational Complexity (CCC)*, 1998, pp. 8–12.

[11] R. Santhanam, "Circuit lower bounds for Merlin-Arthur classes," *SIAM J. Comput.*, vol. 39, no. 3, pp. 1038–1061, 2009.

[12] L. Chen and R. Tell, "Bootstrapping results for threshold circuits "just beyond" known lower bounds," in *Proc. 51st Annual ACM Symposium on Theory of Computing (STOC)*, 2019, pp. 34–41.

[13] D. M. Kane and R. Williams, "Super-linear gate and super-quadratic wire lower bounds for depth-two and depth-three threshold circuits," in *Proc. 48th Annual ACM Symposium on Theory of Computing (STOC)*, 2016, pp. 633–643.

[14] R. Williams, "Improving exhaustive search implies superpolynomial lower bounds," *SIAM J. Comput.*, vol. 42, no. 3, pp. 1218–1244, 2013.

[15] E. Ben-Sasson and E. Viola, "Short PCPs with projection queries," in *Proc. 41st International Colloquium on Automata, Languages and Programming (ICALP)*, 2014, pp. 163–173.

[16] C. Murray and R. Williams, "Circuit lower bounds for nondeterministic quasi-polytime: An easy witness lemma for np and nqp," in *Proc. 50th Annual ACM Symposium on Theory of Computing (STOC)*, 2018.

[17] L. Chen, X. Lyu, and R. Williams, "Almost-everywhere circuit lower bounds from non-trivial derandomization," in *Proc. 61st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2020.

[18] R. Williams, "Non-uniform ACC circuit lower bounds," in *Proc. 26th Annual IEEE Conference on Computational Complexity (CCC)*, 2011, pp. 115–125.

[19] S. Aaronson, "$P \stackrel{?}{=} NP$," in *Open Problems in Mathematics*, J. F. Nash, Jr. and M. T. Rassias, Eds. Springer International Publishing, 2016, pp. 1–122.

[20] R. Santhanam and R. Williams, "On medium-uniformity and circuit lower bounds," in *Proc. 28th Annual IEEE Conference on Computational Complexity (CCC)*, 2013, pp. 15–23.

[21] L. Chen and R. R. Williams, "Stronger Connections Between Circuit Analysis and Circuit Lower Bounds, via PCPs of Proximity," in *Proc. 34th Annual IEEE Conference on Computational Complexity (CCC)*, 2019, pp. 19:1–19:43.

[22] P. Gopalan, D. M. Kane, and R. Meka, "Pseudorandomness via the discrete Fourier transform," *SIAM J. Comput.*, vol. 47, no. 6, pp. 2451–2487, 2018.

[23] I. Diakonikolas, P. Gopalan, R. Jaiswal, R. A. Servedio, and E. Viola, "Bounded independence fools halfspaces," *SIAM J. Comput.*, vol. 39, no. 8, pp. 3441–3462, 2010.

[24] Y. Rabani and A. Shpilka, "Explicit construction of a small ϵ-net for linear threshold functions," *SIAM J. Comput.*, vol. 39, no. 8, pp. 3501–3520, 2010.

[25] I. Diakonikolas, D. M. Kane, and J. Nelson, "Bounded independence fools degree-2 threshold functions," in *Proc. 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2010, pp. 11–20.

[26] D. M. Kane, "A small PRG for polynomial threshold functions of Gaussians," in *Proc. 52nd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2011, pp. 257–266.

[27] Z. S. Karnin, Y. Rabani, and A. Shpilka, "Explicit dimension reduction and its applications," *SIAM J. Comput.*, vol. 41, no. 1, pp. 219–249, 2012.

[28] R. Meka and D. Zuckerman, "Pseudorandom generators for polynomial threshold functions," *SIAM J. Comput.*, vol. 42, no. 3, pp. 1275–1301, 2013.

[29] D. M. Kane, "A pseudorandom generator for polynomial threshold functions of Gaussian with subpolynomial seed length," in *Proc. 29th Annual IEEE Conference on Computational Complexity (CCC)*, 2014, pp. 217–228.

[30] P. K. Kothari and R. Meka, "Almost optimal pseudorandom generators for spherical caps," in *Proc. 47th Annual ACM Symposium on Theory of Computing (STOC)*, 2015, pp. 247–256.

[31] P. Gopalan, R. O'Donnell, Y. Wu, and D. Zuckerman, "Fooling functions of halfspaces under product distributions," in *Proc. 25th Annual IEEE Conference on Computational Complexity (CCC)*, 2010, pp. 223–234.

[32] P. Harsha, A. Klivans, and R. Meka, "An invariance principle for polytopes," *J. ACM*, vol. 59, no. 6, pp. 29:1–29:25, 2012.

[33] R. A. Servedio and L.-Y. Tan, "Fooling intersections of low-weight halfspaces," in *Proc. 58th Annual IEEE Conference on Computational Complexity (CCC)*, 2017, pp. 824–835.

[34] E. Chattopadhyay, A. De, and R. A. Servedio, "Simple and efficient pseudorandom generators from Gaussian processes," in *Proc. 34th Annual IEEE Conference on Computational Complexity (CCC)*, 2019, pp. Art. No. 4, 33.

[35] R. O'Donnell, R. A. Servedio, and L.-Y. Tan, "Fooling polytopes," in *Proc. 51st Annual ACM Symposium on Theory of Computing (STOC)*, 2019, pp. 614–625.

[36] V. Kabanets, S. Koroth, Z. Lu, D. Myrisiotis, and I. C. Oliveira, "Algorithms and lower bounds for de morgan formulas of low-communication leaf gates," in *Proc. 35th Annual IEEE Conference on Computational Complexity (CCC)*, 2020.

[37] J. Alman, T. M. Chan, and R. Williams, "Polynomial representations of threshold functions and algorithmic applications," in *Proc. 57th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2016, pp. 467–476.

[38] R. Impagliazzo, R. Paturi, and S. Schneider, "A satisfiability algorithm for sparse depth two threshold circuits," in *Proc. 54th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2013, pp. 479–488.

[39] R. R. Williams, "New algorithms and lower bounds for circuits with linear threshold gates," *Theory Comput.*, vol. 14, pp. 1–25 (Art. 17), 2018.

[40] S. Tamaki, "A satisfiability algorithm for depth two circuits with a sub-quadratic number of symmetric and threshold gates," Electronic Colloquium on Computational Complexity: ECCC, Tech. Rep. TR16-100, 2016.

[41] R. Tell, "Quantified derandomization of linear threshold circuits," in *Proc. 50th Annual ACM Symposium on Theory of Computing (STOC)*, 2018, pp. 855–865.

[42] P. Hatami, W. M. Hoza, A. Tal, and R. Tell, "Fooling constant-depth threshold circuits," Electronic Colloquium on Computational Complexity: ECCC, Tech. Rep. TR21-002, 2021.

[43] H. D. Gröger and G. Turán, "On linear decision trees computing boolean functions," in *Proc. 18th International Colloquium on Automata, Languages and Programming (ICALP)*, 1991, pp. 707–718.

[44] V. P. Roychowdhury, A. Orlitsky, and K.-Y. Siu, "Lower bounds on threshold and related circuits via communication complexity," *IEEE Trans. Inf. Theory*, vol. 40, no. 2, pp. 467–474, 1994.

[45] N. Nisan, "The communication complexity of threshold gates," in *Combinatorics, Paul Erdős is eighty, Vol. 1*, ser. Bolyai Society Mathematical Studies, 1993, pp. 301–315.

[46] V. Kabanets and J.-Y. Cai, "Circuit minimization problem," in *Proc. 32nd Annual ACM Symposium on Theory of Computing (STOC)*, 2000, pp. 73–79.

[47] A. A. Razborov and S. Rudich, "Natural proofs," *J. Comput. System Sci.*, vol. 55, no. 1, part 1, pp. 24–35, 1997.

[48] L. Chen, C. Jin, and R. R. Williams, "Hardness magnification for all sparse NP languages," in *Proc. 60th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2019.

[49] I. C. Oliveira and R. Santhanam, "Hardness magnification for natural problems," in *Proc. 59th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2018, pp. 65–76.

[50] I. C. Oliveira, J. Pich, and R. Santhanam, "Hardness magnification near state-of-the-art lower bounds," in *Proc. 34th Annual IEEE Conference on Computational Complexity (CCC)*, 2019, pp. Art. No. 27, 29.

[51] L. Chen, D. M. McKay, C. D. Murray, and R. R. Williams, "Relations and Equivalences Between Circuit Lower Bounds and Karp-Lipton Theorems," in *Proc. 34th Annual IEEE Conference on Computational Complexity (CCC)*, 2019, pp. 30:1–30:21.

[52] M. Cheraghchi, V. Kabanets, Z. Lu, and D. Myrisiotis, "Circuit lower bounds for MCSP from local pseudorandom generators," in *Proc. 46th International Colloquium on Automata, Languages and Programming (ICALP)*, vol. 132, 2019, pp. Art. No. 39, 14.

[53] A. Golovnev, R. Ilango, R. Impagliazzo, V. Kabanets, A. Kolokolova, and A. Tal, "$\mathrm{AC}^0[p]$ lower bounds against MCSP via the coin problem," in *Proc. 46th International Colloquium on Automata, Languages and Programming (ICALP)*, vol. 132, 2019, pp. Art. No. 66, 15.

[54] J. Håstad, "The shrinkage exponent of De Morgan formulas is 2," *SIAM J. Comput.*, vol. 27, no. 1, pp. 48–64, 1998.

[55] B. A. Subbotovskaja, "Realization of linear functions by formulas using $\vee$, $\&$, $-$," *Soviet Mathematics. Doklady*, vol. 2, pp. 110–112, 1961.

[56] V. M. Khrapčenko, "A certain method of obtaining estimates from below of the complexity of π-schemes," *Matematicheskie Zametki*, vol. 10, pp. 83–92, 1971.

[57] A. E. Andreev, "On a method for obtaining more than quadratic effective lower bounds for the complexity of π-schemes," *Vestnik Moskovskogo Universiteta. Seriya I. Matematika, Mekhanika*, no. 1, pp. 70–73, 103, 1987.

[58] R. Impagliazzo and N. Nisan, "The effect of random restrictions on formula size," *Random Structures & Algorithms*, vol. 4, no. 2, pp. 121–133, 1993.

[59] M. S. Paterson and U. Zwick, "Shrinkage of De Morgan formulae under restriction," *Random Structures & Algorithms*, vol. 4, no. 2, pp. 135–150, 1993.

[60] A. Tal, "Shrinkage of De Morgan formulae by spectral techniques," in *55th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2014*. IEEE Computer Soc., Los Alamitos, CA, 2014, pp. 551–560.

[61] ——, "Formula lower bounds via the quantum method," in *Proc. 49th Annual ACM Symposium on Theory of Computing (STOC)*, 2017, pp. 1256–1268.

[62] I. Komargodski, R. Raz, and A. Tal, "Improved average-case lower bounds for De Morgan formula size: matching worst-case lower bound," *SIAM J. Comput.*, vol. 46, no. 1, pp. 37–57, 2017.

[63] A. Bogdanov, "Small bias requires large formulas," in *Proc. 45th International Colloquium on Automata, Languages and Programming (ICALP)*, vol. 107, 2018, pp. 22:1–22:12.

[64] R. Santhanam, "Fighting perebor: new and improved algorithms for formula and QBF satisfiability," in *Proc. 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2010, pp. 183–192.

[65] I. Komargodski and R. Raz, "Average-case lower bounds for formula size," in *Proc. 45th Annual ACM Symposium on Theory of Computing (STOC)*, 2013, pp. 171–180.

[66] R. Impagliazzo and V. Kabanets, "Fourier concentration from shrinkage," *Comput. Complexity*, vol. 26, no. 1, pp. 275–321, 2017.

[67] M. Ajtai and A. Wigderson, "Deterministic simulation of probabilistic constant depth circuits," in *Proc. 26th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 1985.

[68] P. Gopalan, R. Meka, O. Reingold, L. Trevisan, and S. Vadhan, "Better pseudorandom generators from milder pseudorandom restrictions," in *Proc. 53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2012, pp. 120–129.

[69] E. Chattopadhyay, P. Hatami, K. Hosseini, and S. Lovett, "Pseudorandom generators from polarizing random walks," *Theory Comput.*, vol. 15, pp. Paper No. 10, 26, 2019.

[70] J. Håstad, *Computational Limitations of Small-depth Circuits*. MIT Press, 1987.

[71] R. Chen, V. Kabanets, A. Kolokolova, R. Shaltiel, and D. Zuckerman, "Mining circuit lower bound proofs for meta-algorithms," *Comput. Complexity*, vol. 24, no. 2, pp. 333–392, 2015.

[72] R. Impagliazzo, R. Paturi, and F. Zane, "Which problems have strongly exponential complexity?" *J. Comput. System Sci.*, vol. 63, no. 4, pp. 512–530, 2001.

[73] D. Gavinsky, S. Lovett, M. Saks, and S. Srinivasan, "A tail bound for read-k families of functions," *Random Structures & Algorithms*, vol. 47, no. 1, pp. 99–108, 2015.

[74] A. A. Sherstov, "Making polynomials robust to noise," *Theory Comput.*, vol. 9, pp. 593–615, 2013.

[75] V. Kabanets and Z. Lu, "Satisfiability and derandomization for small polynomial threshold circuits," in *Proc. 22nd International Workshop on Randomization and Approximation Techniques in Computer Science (RANDOM)*, 2018, pp. Art. No. 46, 19.

[76] R. A. Servedio and L.-Y. Tan, "Luby-Veličković-Wigderson revisited: improved correlation bounds and pseudorandom generators for depth-two circuits," in *Proc. 22nd International Workshop on Randomization and Approximation Techniques in Computer Science (RANDOM)*, vol. 116, 2018, pp. Art. No. 56, 20.

[77] E. Viola, "Pseudorandom bits for constant-depth circuits with few arbitrary symmetric gates," *SIAM J. Comput.*, vol. 36, no. 5, pp. 1387–1403, 2007.

[78] S. Lovett and S. Srinivasan, "Correlation bounds for poly-size AC^0 circuits with $n^{1-o(1)}$ symmetric gates," in *Proc. 14th International Workshop on Randomization and Approximation Techniques in Computer Science (RANDOM)*, 2011, pp. 640–651.

[79] T. Sakai, K. Seto, S. Tamaki, and J. Teruyama, "Bounded depth circuits with weighted symmetric gates: satisfiability, lower bounds and compression," in *Proc. 41st International Symposium on Mathematical Foundations of Computer Science*, 2016.