Fooling Constant-Depth Threshold Circuits

Pooya Hatami * William M. Hoza † Avishay Tal ‡ Roei Tell §

December 7, 2022

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 $\epsilon = 2^{-n/\mathrm{polylog}(n)}$.

*Department of Computer Science and Engineering, The Ohio State University, OH, USA. Email: pooyahat@gmail.com

†Department of Computer Science, University of Texas at Austin, TX, USA. Email: whoza@utexas.edu

‡Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, CA, USA. Email: atal@berkeley.edu

§Massachusetts Institute of Technology, Cambridge, MA. Email: roei.tell@gmail.com

Contents

1 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 [IMZ19]. 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 [Hås14; CSS18; ST17a; Tal17b], 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.

1.1 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. 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 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 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 [IPS93], who showed that the parity function is hard for LTF circuits of depth

[1] The class TC^0 is sometimes defined using unweighted majority gates and sometimes defined using LTF gates. Both definitions are equivalent up to polynomial overheads (see [GHR92; GK98]), but since we will be concerned with precise size bounds we will use a specific definition. Note that our PRG fools the stronger class.

d with $n^{1+c^{-d}}$ wires, for some constant $c > 1$.[2] Despite the fact that no lower bounds for larger 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 [CSS18]. Also, for the special case of $d = 2$, Kane and Williams [KW16] 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. [Wil13; BSV14; MW18; CLW20]) 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 CAPP[3] that are faster than the trivial brute-force algorithm. Following Williams' [Wil11] 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 LTF circuits by constructing circuit-analysis algorithms for such circuits – see, e.g., the first open problem in [Aar16], and also see [Wil13; SW13; MW18; CW19]. 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 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 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 [GKM18], following [DGJSV10; RS10; DKN10; Kan11; KRS12; MZ13; Kan14; KM15]). Various works constructed PRGs for "simple functions" of LTFs, for example for $\mathsf{AND} \circ \mathsf{LTF}$ (aka polytopes, see [GOWZ10; DKN10; HKM12; ST17b; CDS19; OST19; KKLMO20]). For LTF circuits of *depth two* and subquadratic size, a PRG with seed length $n^{1-\Omega(1)}$ was constructed by Servedio and Tan [ST17a]; and a satisfiability algorithm with running time $2^{n-n^{\Omega(1)}}$ was constructed by Alman, Chan, and Williams [ACW16] (following [IPS13; Wil18; Tam16]; this algorithm also works for the larger class $\mathsf{AC}^0[m] \circ \mathsf{LTF}_{n^{2-\Omega(1)}} \circ \mathsf{LTF}$, see Section 3). However, despite these efforts, for circuits of *arbitrary depth* $d > 2$, prior to this work no non-trivial deterministic satisfiability or CAPP algorithm was known. The only known deterministic circuit-analysis algorithm for such circuits was an algorithm for the relaxed version of CAPP called *quantified derandomization* (see [Tel18]), algorithms for which are not known to imply lower bounds. We defer further discussion of other relevant works to Section 3.

Building on the rich ideas developed in the last decade, in this paper we construct the *first non-trivial PRG for* 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 [IPS93] and the average-case lower bounds of [CSS18]. Thus, our construction brings the extensive research effort described above to the point where essentially further improvement would yield new lower bounds for LTF circuits.

Theorem 1.1 (PRG for super-linear 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* 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 1.1 holds also for super-constant values of $d \in \mathbb{N}$ (see Theorem 6.25 for details). Parsing the parameters of our PRG, the seed length $O(n^{1-\delta})$ is

[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 LTF circuits) against functions that are "not very explicit", and in particular are not even known to be in NP (see, e.g., [Kan82; BFT98; San09]). We also note that the precise value of the constant $c > 1$ in this expression turns out to be surprisingly important (see [CT19]).

[3]Recall that 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 $2/3$.

"slightly non-trivial" (yielding a 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 LTF circuits (see Proposition 4.11). 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 LTF circuits (with respect to a natural polynomial-time-samplable distribution; again, see Proposition 4.11). 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 elaboare on this in Section 2.1).

As part of our proof of Theorem 1.1 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 1.1. 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 3 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 1.2 (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* $\widetilde{O}\left(\sqrt{n \cdot (s + \log(1/\epsilon))}\right)$.

One corollary of Theorem 1.2 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* (see 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 [GT91; ROS94; Nis93]).

A stronger efficiency requirement and a new lower bound. The PRG in Theorem 1.1 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 $O(|s|) = O(n^{1-\delta})$ (see Theorem 6.25).

The existence of PRGs meeting such a strong efficiency requirement implies that the fooled circuit class cannot solve the Minimum Circuit Size Problem (MCSP) [KC00]. 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 [KC00; RR97].) 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 1.3 (MCSP lower bound for LTF circuits of super-linear size; see 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 1.3 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 1.3 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 [CJW19], which imply that for some constant $c > 1$, if for all $\beta > 0$

it holds that 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 [CT19] (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 [OS18; OPS19; CJW19; CMMW19; CKLM19; GII+19]).

1.2 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 [Hås98] (following [Sub61; Khr71; And87; IN93; PZ93]), with subsequent log-factor improvements [Tal14; Tal17a]. These bounds were extended to average-case lower bounds by Komargodski, Raz, and Tal [KRT17] and Bogdanov [Bog18] (following [San10; KR13]; see also [IK17; Tal17a]), 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 [IMZ19] 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 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 [KKLMO20, 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 1.4 (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

[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 [CJW19, Theorem 1.1, Item 7]. Thus, intuitively, the difference between the unconditional result in Corollary 1.3 and a result that would imply lower bounds for all of TC^0 is even smaller.

[5] More precisely, the result statements in [IMZ19] 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.

than $1/2 + 2^{n^{-\delta}}$ over a natural polynomial-time-samplable distribution, for any $\delta > 0$ (see Proposition 4.11). This essentially matches the best known average-case lower bounds for De Morgan formulas by [KRT17; Bog18], which were mentioned above.

2 High-level proof overviews

We now present high-level overviews of our proofs. First, in Section 2.1, we will describe the common high-level technical challenge underlying both constructions, and our general approach for handling this challenge. Then in Section 2.2 we describe our construction of a PRG for De Morgan formulas (i.e., Theorem 1.4), which is considerably simpler than our PRG for LTF circuits and nevertheless showcases our approach. Finally, in Section 2.3, we move on to the more involved construction of a PRG for LTF circuits (Theorem 1.1).

2.1 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.[6] There are many known frameworks for obtaining PRGs from pseudorandom restrictions (see, e.g. [AW85; IMZ19; GMRTV12; CHHL19]), 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 [Hås87; Hås98]), 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.[7]

Our way to bypass this obstacle in both settings, generalizing ideas from [Hås14; CSS18; ST17a; Tal17b], 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 [CSS18], we will show that with probability $1 - \exp(-n^{\Omega(1)})$, restricted LTF circuits of super-linear size can be approximated

[6] Recall that a restriction is a partial assignment to the input variables. Following standard convention, throughout the section the letter p will denote the probability that each variable remains alive (i.e., unassigned) under a random or pseudorandom restriction.

[7] In fact, to materialize our approach and get a non-trivial PRG for LTF circuits we will have to show restrictions that fail with *exponentially small* probability (which is indeed what we show); see Section 2.3 for details.

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 2.2 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 that a suitable *pseudorandom restriction* also simplifies the circuit to the corresponding hybrid model with probability $1 - \exp(-n^{\Omega(1)})$.[8] 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 1.2), whereas for De Morgan formulas we will refine an extractor-based argument of [IMZ19] 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 [IMZ19]. 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 [IMZ19]. However, for these two classes, it turns out that a more elementary approach gives even better parameters. We defer the details to Appendix B.

2.2 Low-error PRG for De Morgan formulas

For De Morgan formulas we will build on the PRG framework of Impagliazzo, Meka, and Zuckerman [IMZ19], 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 [Hås98; Tal14]), and in [IMZ19] 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 [IMZ19, Lemma 4.8]).

As mentioned in Section 2.1, 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.[9] 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:

[8] 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).

[9] Recall that (by a counting argument) for any $c \in \mathbb{N}$ there exist functions over $O(\log(n))$ variables that require formulas of size n^c, so this counter-example holds also for polynomial-sized formulas.

Proposition 2.1 (low-error pseudorandom restrictions for De Morgan formulas; informal, see Theorem 5.7)**.** *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 $f|_\rho$ 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.*[10]

Let us first describe the main idea in the proof of Proposition 2.1. Recall that a De Morgan formula is called read-k if each variable appears at most k times among the leaves. In [IMZ19] 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 [KR13; KRT17; CKKSZ15] 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 2.1, we follow [IMZ19] 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 2.1.

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 [IMZ19] 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 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 $\log(m/\epsilon)$. We defer the full description to Section 5.4.[11]

2.3 PRG for super-linear LTF circuits

We now describe the proof of Theorem 1.1. 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^δ}. As part of this proof we will also describe the proof of Theorem 1.2 (our PRG for $\mathtt{ANY}_s \circ \mathtt{LTF}_n$), and a self-contained description of the latter appears in Section 2.3.3.

[10] 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 (see Section 5 for details).

[11] In a nutshell, instead of treating the entire tree as a single computational device with exponential description length, we express the tree as a sum over m functions, one per leaf. Then, to ϵ-fool the entire tree it suffices to ϵ/m-fool every single path in the tree. Indeed, each path has a succinct description that fits the PRG framework of [IMZ19], and the crucial point for us is that in this framework the seed length only increases *logarithmically* in the error ϵ/m of *"fooling" the restricted model*. See Section 5.4 for further details.

2.3.1 Overview: Basic ideas and main challenges

For this setting we will use the classical restrictions-to-PRG framework of Ajtai and Wigderson [AW85]. 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 [IPZ01; CSS18]. In the most relevant result to our setting, Chen, Santhanam, and Srinivasan [CSS18, 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 [Tel18]). 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 [AW85] 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 [CSS18; Tel18] 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".[12] 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 2.1, the [AW85] 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 [Tel18] only bounds the error by $p^{1/5}$, using a naive concentration bound (i.e., Markov's inequality); whereas the analysis of [CSS18] for truly uniform restrictions relies on a read-k Chernoff bound (i.e., on [GLSS15]), 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 2.3.3, previously-known PRG constructions do not seem to suffice for this purpose.

We now state our two key technical results underlying Theorem 1.1, 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* LTFs *and variables*, with no more than $p^{\Omega(1)} \cdot (p \cdot n)$

[12] In [Tel18], 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 [CSS18].

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 2.2 (low-error pseudorandom restrictions for super-linear LTF circuits; informal, see Corollary 6.12)**.** *For any constant $d \in \mathbb{N}$ and $\delta = \frac{1}{2} \cdot 50^{-d}$, there is a distribution over subsets $\mathbf{I} \subseteq [n]$ 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 $[n] \setminus \mathbf{I}$, 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 2.2, 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 1.2.

Proposition 2.3 (low-error PRG for the hybrid model; informal, see Theorem 6.21)**.** *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 2.2, 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 2.3 will be dominated by $\widetilde{O}(\sqrt{(p \cdot n) \cdot D}) \leq p^{\Omega(1)} \cdot (p \cdot n)$.

In the following Sections 2.3.2 and 2.3.3 we will describe the main ideas behind the proofs of Propositions 2.2 and 2.3, respectively. We note that these two sections can be read independently of each other.

2.3.2 Low-error pseudorandom restrictions that "sufficiently simplify" the circuit

We now describe the proof of Proposition 2.2, 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 [CSS18; Tel18], 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). See Claim 6.7.1 for precise details.

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 [CSS18] for a truly random restriction handled this subgraph with an exponentially small failure probability; but this analysis relied on a read-k Chernoff bound [GLSS15], 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 (see Section 6.1.2). 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. For further details see the proof of Proposition 6.7.

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. (As mentioned above, we actually construct both an *upper-sandwiching* approximation and a *lower-sandwiching* approximation, at the cost of labeling each leaf by one depth-$(d-1)$ LTF circuit and a small set of LTF functions. For simplicity, we ignore this complication in the current overview.)

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 $\epsilon_{i-1} \cdot 2^{M_i} \ll \epsilon_i$. For further details see the proof of Proposition 6.9.

2.3.3 Low-error PRG for the "sufficiently simple" hybrid model

Our goal in this section is to prove Theorem 1.2, i.e., to construct a PRG for the class $\mathtt{ANY}_s \circ \mathtt{LTF}_n$ of functions that can be computed as an arbitrary function of s LTFs, whose seed length is $\widetilde{O}\left(\sqrt{n \cdot (s + \log(1/\epsilon))}\right)$. Our main application of this result is to prove Proposition 2.3, and we will explain at the end of this section how the latter can be easily obtained as a corollary. 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 $\mathtt{AND} \circ \mathtt{LTF}$, was proportional to $\log(1/\epsilon)^2$, which is too much for us (see [GOWZ10; HKM12; ST17b;

OST19]). However, very recently Kabanets, Koroth, Lu, Myrisiotis, and Oliveira [KKLMO20] 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 $\widetilde{O}\left(\sqrt{n} \cdot s^{1/4} \cdot \log(1/\epsilon)\right)$. 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 $\texttt{LTF}$s and for some composition function h. Informally, the main idea underlying [KKLMO20] 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 [KKLMO20, Theorem 25] for a detailed analysis.)

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 $\widetilde{O}\left(\sqrt{n \cdot s} \cdot \log(1/\epsilon)\right)$ (see Section 6.2 for details). 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 [She13]: 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 $\left|p_h(x) - \widetilde{p_h}(x+\eta)\right| < \epsilon$. 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 [She13], 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

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

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 $\texttt{LTF}$s) and PRGs (for the communication protocols) as above, and obtain a PRG for $\texttt{ANY}_s \circ \texttt{LTF}$ with seed length $\widetilde{O}(\sqrt{n \cdot d} \cdot \log(1/\rho)) = \widetilde{O}(\sqrt{n \cdot (s + \log(1/\epsilon))})$.

As an aside, observe that our first step was to use a trivial polynomial of degree s to compute the composition h. This is unavoidable when h is arbitrary, but when h comes from

more a restricted class we can use polynomials with better parameters, yielding a PRG with shorter seed. See Appendix A for details.

Proposition 2.3 as a corollary of Theorem 1.2. Let T be a DT with parameters as in Proposition 2.3. We can compute T as $T(x) = \sum_{\ell \in L} I_\ell(x) \cdot \Phi_\ell(x)$, where the summation is over $|L| \leq 2^{D+M}$ leaves, and for each leaf ℓ the function I_ℓ is the indicator function of ℓ, and $\Phi_\ell(x)$ is the LTF function that labels ℓ.[13] It follows that to fool T with error δ, it suffices to fool each term $I_\ell \cdot \Phi_\ell$ with error $\epsilon = \delta \cdot 2^{-(D+M)}$. Now, since the path to each leaf queries D variables and M LTFs, we can express I_ℓ as a conjunction of $M+1$ LTFs, one of which will simultaneously test the values of all the D queried variables. Therefore, $I_\ell \cdot \Phi_\ell$ is a conjunction of $M+2$ LTFs, so we can apply Theorem 1.2 with the specific composition function $h = \text{AND}$.

2.3.4 Making our PRG strongly explicit

The description above only claims that the PRG is computable in polynomial time. However, as mentioned in Section 1.1, our PRG is in fact *strongly explicit*, in the sense that given a seed s and an index $i \in [n]$ of an output, we can compute the corresponding output bit $G(s)_i$ in time $O(n^{1-\delta})$. The reason is that our algorithmic construction essentially just combines a large number of k-wise independent distributions and PRGs for communication protocols, with varying parameters, and strongly explicit constructions for both objects are known. To ensure that even the *combination* of these objects is strongly explicit – that is, the combination preserves this property – we fool communication protocols using a recent PRG by Forbes and Kelley [FK18], which is algorithmically simple and thus suited for our purposes. See Section 6.2.1 for an explanation and for technical details.

3 Previous work on circuit-analysis algorithms for LTF circuits

As mentioned in Section 1.1, a large number of previous works focused on circuit-analysis algorithms for LTF circuits. We now survey the previously-known results, while focusing on *deterministic* circuit-analysis algorithms. (Many randomized circuit-analysis algorithms for LTF circuits are known – see, e.g., [CSS18; ACW16; KL18; KKLMO20] and the references therein – but these are not the focus of this work, and do not imply circuit lower bounds via Williams' [Wil13] approach.)

Single LTFs and simple compositions of LTFs. For a single LTF function, a PRG with near-optimal seed length $\widetilde{O}(\log(n/\epsilon))$ was constructed by Gopalan, Kane, and Meka [GKM18], following [DGJSV10; RS10; DKN10; Kan11; KRS12; MZ13; Kan14; KM15]. Concurrently and subsequently, various PRGs were constructed for "simple compositions" of LTFs, and in particular for $\text{AND} \circ \text{LTF}$ (i.e., for polytopes, see [GOWZ10; DKN10; HKM12; ST17b; CDS19; OST19; KKLMO20]), for monotone functions of LTFs [GOWZ10], and for small De Morgan formulas of LTFs [KKLMO20].

The class $\text{ANY}_{o(n)} \circ \text{LTF}$. The problem of fooling $\text{ANY}_s \circ \text{LTF}$ with error ϵ reduces to fooling $\text{AND}_s \circ \text{LTF}$ with error $\epsilon/2^s$ [CDS19, Footnote 1]. Combining this reduction with the PRG of [KKLMO20, Theorem 30], one can obtain a PRG for $\text{ANY}_s \circ \text{LTF}$ with seed length $\widetilde{O}(\sqrt{n} \cdot (s^{5/4} + s^{1/4} \cdot \log(1/\epsilon)))$, which is non-trivial for $s \leq n^{2/5}/\text{polylog}(n)$ (note that this result is superseded by Theorem 1.2). Chattopadhyay, De, and Servedio [CDS19]

[13] Recall that in our actual hybrid model, each leaf is not labeled by a single LTF but rather by a small set of $\text{poly}(n)$ LTFs. We again ignore this issue in the overview for simplicity, and in the actual proof we will simply union-bound over the $\text{poly}(n)$ LTFs at each leaf (see Theorem 6.21 for details).

(following [GOWZ10]) 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 $s \geq n^{\Omega(1)}$ or $\epsilon \leq n^{-\Omega(1)}$. In comparison, the PRG from Theorem 1.2 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 $\mathtt{LTF}$ circuits. For circuits of depth two (i.e., $\mathtt{LTF} \circ \mathtt{LTF}$ circuits), Servedio and Tan [ST17a] 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 [ACW16] (following [IPS13; Wil18; Tam16]) 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 $\mathtt{LTF}$ gates at their bottom layer.

For $\mathtt{LTF}$ 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 $\mathtt{LTF}$ gates: Specifically, Servedio and Tan [ST18] (following [Vio07]) constructed a PRG for AC^0 circuits of size S with at most $2^{\alpha \cdot \sqrt{\log S}}$ $\mathtt{LTF}$ gates (for a universal constant $\alpha > 0$) whose seed length is $2^{O(\sqrt{\log S})} + \mathrm{polylog}(1/\epsilon)$; and Lovett and Srinivasan [LS11] constructed an incomparable PRG for AC^0 circuits of polynomial size with at most n^{49} $\mathtt{LTF}$ gates whose seed length is n^{δ} (for an arbitrarily small $\delta > 0$) and whose error is $2^{-n^{24}}$. See [SSTT16] for another result in this spirit.

The only previously-known algorithm for $\mathtt{LTF}$ 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 [Tel18]. In general, quantified derandomization algorithms are not known to imply lower bounds; however, even a very mild improvement in the number of exceptional inputs that the foregoing algorithm can handle (namely, an improvement in the universal constant hiding in the O-notation) would yield new lower bounds for TC^0 [CT19].

4 Preliminaries

We denote distributions and random variables in boldface. We let $\mathbf{u}_n$ denote the uniform distribution over $\{0,1\}^n$. We will also use the convention of first defining some parameter values (e.g., an input length n and auxiliary parameters denoting size and error bounds) and then asserting that a uniform algorithm runs within some time bound with respect to these parameter values; in all such statements, we assume that the algorithm explicitly gets all the relevant parameter values as part of its input.

When we bound the running time of computing *each output symbol* of a function $G\colon \{0,1\}^{\ell} \to \Sigma^n$, the intention is that the algorithm also gets an index $i \in [n]$ as part of its input and its job is to output the i-th symbol of the output of G. When we say that a distribution over strings $\mathbf{x} \in \Sigma^n$ can be sampled using s truly random bits such that *each coordinate of* $\mathbf{x}$ *can be computed in time* t, we mean that there is a generator $G\colon \{0,1\}^s \to \Sigma^n$ such that $G(\mathbf{u}_s)$ is distributed according to $\mathbf{x}$ and each output symbol of G can be computed in time t. As a special case, when we say that a distribution over sets $\mathbf{I} \subseteq [n]$ can be sampled using s truly random bits such that *membership in* $\mathbf{I}$ *can be computed in time* t, we mean that each coordinate of the indicator vector of $\mathbf{I}$ can be computed in time t.

In the end, we will claim that each output bit of our PRG for constant-depth $\mathtt{LTF}$ circuits can be computed in time $n^{1-\Omega(1)}$. Fine-grained time complexity can in general be sensitive

to polynomial overheads caused by different machine models, but our results hold in most reasonable computational models. Indeed, we will prove that the time bounds hold with respect to the standard *multitape Turing machine* model, which is a relatively weak model of computation when we are concerned about polynomial factors. The reason we choose this model is that multitape Turing machines can be implemented by Boolean circuits with very little overhead [PF79], which will facilitate our proof that MCSP cannot be computed by constant-depth slightly-superlinear-size LTF circuits (i.e., the proof of Corollary 1.3).

Definition 4.1 (distance between functions)**.** *We say that two functions* $f, g\colon \{0,1\}^n \to \{0,1\}$ *are* ϵ-close *if* $|\mathbb{E}[f(\mathbf{u}_n)] - \mathbb{E}[g(\mathbf{u}_n)]| \leq \epsilon$.

Definition 4.2 (sandwiching)**.** *We say that* $f\colon \{0,1\}^n \to \{0,1\}$ *is* ϵ-upper-sandwiched *by* $g\colon \{0,1\}^n \to \mathbb{R}$ *if for every* $x \in \{0,1\}^n$ *it holds that* $f(x) \leq g(x)$*, and also* $\mathbb{E}[g(\mathbf{u}_n) - f(\mathbf{u}_n)] \leq \epsilon$*. Similarly, we say that* f *is* ϵ-lower-sandwiched *by* g *if for every* x *it holds that* $g(x) \leq f(x)$ *and we have that* $\mathbb{E}[f(\mathbf{u}_n) - g(\mathbf{u}_n)] \leq \epsilon$.

4.1 Standard concentration bounds

We now state several well-known concentration bounds that we will use in the paper. The first is the standard Hoeffding inequality.

Theorem 4.3 (Hoeffding's inequality [Hoe63, Thm 2])**.** *Let* $w \in \mathbb{R}^n$*, and let* $\mathbf{z}$ *be a uniformly-chosen random vector in* $\{-1,1\}^n$*. Then, for any* $t > 0$ *it holds that*

$$\Pr\left[\left|\langle w, \mathbf{z}\rangle\right| \geq t \cdot \|w\|_2\right] \leq 2\exp(-t^2/2)\,.$$

We also need concentration bounds for two regimes when the random variables are not independent. The first of those, for the case of "read-k" random variables, was proved by Gavinsky, Lovett, Saks, and Srinivasan [GLSS15]; and the second one is for random variables that are k-wise independent, for which we will use the bound of Bellare and Rompel [BR94].

Theorem 4.4 (read-k Chernoff bound; [GLSS15, Thm 1.1])**.** *Let* $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ *be independent random variables, and let* $\mathbf{y} = \mathbf{y}_1, ..., \mathbf{y}_r$ *be a set of functions* $\mathbf{x} \to \{0,1\}$ *such that each* $\mathbf{x}_i$ *influences at most* k *functions from* $\mathbf{y}$*, where* $k \in \mathbb{N}$*. Let* $\mu = \mathbb{E}[\sum_i \mathbf{y}_i]$*. Then, for any* $\Delta > 0$*,*

$$\Pr\left[\sum_{i=1}^{r} \mathbf{y}_i \geq \mu + \Delta\right] \leq \exp\left(-\frac{2\Delta^2}{rk}\right).$$

Theorem 4.5 (tail bound for k-wise independence; [BR94, Lemma 2.3])**.** *Suppose* $\mathbf{x}_1, \ldots, \mathbf{x}_n$ *are* k*-wise independent random variables taking values in* $[0,1]$*, where* $k \geq 4$*.*[14] *Let* $\mu = \mathbb{E}[\sum_i \mathbf{x}_i]$*, and let* Δ *satisfy* $\Delta \geq \mu/2$ *and* $\Delta \geq 300k$*. Then*

$$\Pr\left[\left|\mu - \sum_i \mathbf{x}_i\right| \geq \Delta\right] \leq 2^{-k}.$$

4.2 Restrictions

Given a function $f\colon \{0,1\}^n \to \{0,1\}$, a restriction of f is a subset $W \subseteq \{0,1\}^n$. We will be interested in restrictions that are subcubes, and such restrictions can be described by a string

[14] For convenience, we allow non-integer values of k. The meaning is that any set of at most k of the variables $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are independent. Bellare and Rompel's lemma [BR94] assumes k is an even integer, but we chose the constant 300 to be large enough that our formulation follows from Bellare and Rompel's lemma by considering $\lfloor k \rfloor$ and $\lfloor k \rfloor - 1$.

$\rho \in \{0,1,\star\}^n$, where the subcube consists of all $x \in \{0,1\}^n$ such that for every $i \in [n]$ for which $\rho_i \neq \star$ it holds that $x_i = \rho_i$. The living variables under ρ are the input bits indexed by the set $\{i \in [n] : \rho_i = \star\}$. We will sometimes describe a restriction by a pair $\rho = (I, z)$, where $I \subseteq [n]$ and $z \in \{0,1\}^{[n]\setminus I}$ or $z \in \{0,1\}^n$, with the interpretation being

$$\rho_i = \begin{cases} \star & \text{if } i \in I \\ z_i & \text{if } i \in [n] \setminus I. \end{cases}$$

The function f restricted to a subcube ρ is denoted by $f\restriction_\rho \colon \{0,1\}^n \to \{0,1\}$ such that $f\restriction_\rho(x) = f(y)$, where for every $i \in [n]$ it holds that $y_i = x_i$ if $\rho_i = \star$ and $y_i = \rho_i$ otherwise. We will also consider the composition of restrictions to subcubes, where a composition $\rho = \rho_1 \circ \rho_2$ yields the restricted function $f\restriction_\rho = \left(f\restriction_{\rho_2}\right)\restriction_{\rho_1}$. Specifically, if $\rho_1 \in \{0,1,\star\}^{\rho_2^{-1}(\star)}$ or $\rho_1 \in \{0,1,\star\}^n$, then

$$(\rho_1 \circ \rho_2)_i = \begin{cases} (\rho_2)_i & \text{if } (\rho_2)_i \neq \star \\ (\rho_1)_i & \text{if } (\rho_2)_i = \star. \end{cases} \tag{1}$$

As a special case, if $x \in \{0,1\}^{\rho^{-1}(\star)}$, then $x \circ \rho$ is the string in $\{0,1\}^n$ obtained by using x to fill in the $\star$ positions of ρ.

Definition 4.6. *A p-*regular *restriction is a random variable $\boldsymbol{\rho} \in \{0,1,\star\}^n$ such that for each i, $\Pr[\boldsymbol{\rho}_i = \star] = p$ and $\Pr[\boldsymbol{\rho}_i = 0] = \Pr[\boldsymbol{\rho}_i = 1] = (1-p)/2$. If additionally the coordinates $\boldsymbol{\rho}_1, \ldots, \boldsymbol{\rho}_n$ are independent, we say that the restriction is* truly random. *If $\boldsymbol{\rho}_1, \ldots, \boldsymbol{\rho}_n$ are k-wise independent, we say the restriction is k-*wise independent.

Definition 4.7. *Let $A \subseteq [n]$. A p-*regular subset *of A is a random variable $\mathbf{I} \subseteq A$ such that for each $i \in A$,*

$$\Pr[i \in \mathbf{I}] = p.$$

*We say the subset is k-*wise independent *if the events $i \in \mathbf{I}$ are k-wise independent.*

The next claim says that a p-regular k-wise independent subset of $[n]$ and a p-regular k-wise independent restriction can each be sampled efficiently using $O(k \cdot \log(n/p))$ truly random bits. Indeed, we will show that the time to compute one coordinate of the restriction on a multitape Turing machine scales linearly with k. The construction is standard, but verifying the running time requires us to be slightly careful.

Claim 4.8. *Let $k, n \in \mathbb{N}$ with $k \leq n$ and let $p \in (0,1)$ be a power of two. A p-regular k-wise independent restriction $\boldsymbol{\rho} \in \{0,1,\star\}^n$ (and a p-regular k-wise independent subset $\mathbf{I} \subseteq [n]$) can be sampled using $O(k \cdot \log(n/p))$ truly random bits such that each coordinate of $\boldsymbol{\rho}$ (and membership in $\mathbf{I}$) can be computed in time $k \cdot \operatorname{polylog}(n/p)$ on a multitape Turing machine.*

Proof. Let $\mathbb{F}$ be a finite field of characteristic 2 with $|\mathbb{F}| \geq n$ and $|\mathbb{F}| \geq q$, where $q = 2/p$. The seed consists of a uniform random vector $\mathbf{f} \in \mathbb{F}^k$, which we think of as the list of coefficients of a polynomial $\mathbf{f} \colon \mathbb{F} \to \mathbb{F}$ of degree at most $k-1$. Evaluate $\mathbf{f}$ at n distinct (fixed) elements of $\mathbb{F}$ and take the first $\log q$ bits of each output value, giving a sequence $\mathbf{y} \in [q]^n$. This sequence is k-wise independent [Vad12, Proposition 3.33] and each coordinate is distributed uniformly over $[q]$. Using Horner's rule, $\mathbf{f}$ can be evaluated using only $O(k)$ field operations. Each field operation can be performed in time $\operatorname{polylog}(|\mathbb{F}|)$ [Sho90], so each coordinate of $\mathbf{y}$ can be computed in time $k \cdot \operatorname{polylog}(nq)$. (Note that according to Horner's rule, we access the coefficients of $\mathbf{f}$ in order from high-degree to low-degree, so this runtime holds even in the multitape Turing machine model.) Our p-regular k-wise independent restriction $\boldsymbol{\rho}$ is given by

$$\boldsymbol{\rho}_i = \begin{cases} 0 & \text{if } \mathbf{y}_i \in \{1, 2, \ldots, q/2-1\} \\ 1 & \text{if } \mathbf{y}_i \in \{q/2, q/2+1, \ldots, q-2\} \\ \star & \text{if } \mathbf{y}_i \in \{q-1, q\}. \end{cases}$$

Our p-regular k-wise independent subset $\mathbf{I} \subseteq [n]$ is given by $\mathbf{I} = \boldsymbol{\rho}^{-1}(\star)$. ■

4.3 Threshold functions

For two vectors $w, x \in \mathbb{R}^n$, we denote by $\langle w, x \rangle = \sum_{i \in [n]} w_i \cdot x_i$ the standard inner-product over the reals. For $h < n$, we denote $w_{>h} = (w_{h+1}, \ldots, w_n) \in \mathbb{R}^{n-h}$.

A linear threshold function (LTF) $\Phi \colon \{0,1\}^n \to \{0,1\}$ is a function of the form $\Phi(x) = 1 \iff \langle x, w \rangle > \theta$, where $w \in \mathbb{R}^n$ is a vector of real "weights", and (the "threshold") $\theta \in \mathbb{R}$ is a real number; we denote $\Phi = (w, \theta)$. The following are standard definitions (see, e.g., [Ser07; DGJSV10]) that refer to structural properties of LTFs.

Definition 4.9 (regularity). *For $\mu > 0$, we say that a vector $w \in \mathbb{R}^n$ is μ-*regular *if for every $i \in [n]$ it holds that $|w_i| \leq \mu \cdot \|w\|_2$. An LTF $\Phi = (w, \theta)$ is μ-regular if w is μ-regular.*

Definition 4.10 (critical index). *When $w \in \mathbb{R}^n$ satisfies $|w_1| \geq |w_2| \geq \cdots \geq |w_n|$, the μ-*critical index *of w is defined as the smallest $h \in [n]$ such that $w_{>h}$ is μ-regular (and $h = \infty$ if no such $h \in [n]$ exists). The critical index of an LTF $\Phi = (w, \theta)$ is the critical index of w', where $w' \in \mathbb{R}^n$ is the vector that is obtained from w by permuting the coordinates in order to have $|w'_1| \geq \cdots \geq |w'_n|$.*

4.4 PRGs yield hard functions

The following proposition (which is a minor modification of [Vio09, Prop. 5]) spells out the standard transformation of PRGs to hard functions:

Proposition 4.11 (from PRG to average case hard function). *Suppose that $G \colon \{0,1\}^\ell \to \{0,1\}^n$ is ϵ-pseudorandom for a class $\mathcal{F}$ of functions $f \colon \{0,1\}^n \to \{0,1\}$ that is closed under permuting the input variables. For $\ell' = \ell + \lceil \log(1/\epsilon) \rceil$, define $h \colon \{0,1\}^{\ell'} \to \{0,1\}$ by $h(x) = 1$ iff there exists y such that x is a prefix of $G(y)$. Let $\mathbf{d} \in \{0,1\}^{\ell'}$ be a random variable that with probability $1/2$ is equal to $\mathbf{u}_{\ell'}$ and with probability $1/2$ is equal to $G(\mathbf{u}_\ell)_{[\ell']}$. Then h is $(1/2 + \epsilon)$-average-case hard for $\mathcal{F}$ with respect to $\mathbf{d}$; that is, for every $f \in \mathcal{F}$ that is sensitive to at most ℓ' input variables,*[15]

$$\Pr[f(\mathbf{d}) = h(\mathbf{d})] \leq \frac{1}{2} + \epsilon. \tag{2}$$

Proof. We have[16]

$$\begin{aligned}
\Pr[f(\mathbf{d}) = h(\mathbf{d})] &= \frac{1}{2} \Pr_{u \sim \mathbf{u}_n}[f(u) = h(u_{[\ell']})] + \frac{1}{2} \Pr_{u \sim \mathbf{u}_\ell}[f(G(u)) = h(G(u)_{[\ell']})] \\
&= \frac{1}{2} \Pr_{u \sim \mathbf{u}_n}[f(u) = h(u_{[\ell']})] + \frac{1}{2} \Pr[f(G(\mathbf{u}_\ell)) = 1] \\
&\leq \frac{1}{2} \left(\Pr[f(\mathbf{u}_n) = 0] + \Pr[h(\mathbf{u}_{\ell'}) = 1]\right) + \frac{1}{2} \Pr[f(G(\mathbf{u}_\ell)) = 1] \\
&\leq \frac{1}{2} \left(\Pr[f(\mathbf{u}_n) = 0] + 2^{\ell - \ell'}\right) + \frac{1}{2} \left(\Pr[f(\mathbf{u}_n) = 1] + \epsilon\right) \\
&\leq \frac{1}{2} + \epsilon,
\end{aligned}$$

where the first inequality holds because $f = h$ implies that either $f = h = 0$ or $f = h = 1$, and the second inequality holds by the definition of h and the fact that G fools f. ■

[15] Technically, the domain of f is $\{0,1\}^n$, whereas $\mathbf{d} \in \{0,1\}^{\ell'}$; the expression $f(\mathbf{d})$ should be interpreted as the output of f when the bits of $\mathbf{d}$ are placed at input locations to which f is sensitive.

[16] Since $\mathcal{F}$ is closed under permuting the input variables, we may assume without loss of generality that the variables to which f is sensitive come before the variables that f ignores, hence $f(\mathbf{d})$ is $f(G(\mathbf{u}_\ell))$ with probability $1/2$ and $f(\mathbf{u}_n)$ otherwise.

5 An improved low-error PRG for De Morgan formulas

Recall that Impagliazzo, Meka, and Zuckerman [IMZ19] designed a PRG for size-s De Morgan formulas ($s \geq n$) with seed length $s^{1/3+o(1)}$ and error $1/\mathrm{poly}(s)$. In this section, we describe a PRG with much lower error. In particular, we fool size-s De Morgan formulas with seed length $s^{1/3+o(1)} \cdot \mathrm{polylog}(n/\epsilon)$.

Theorem 5.1 (Low-error PRG for De Morgan formulas; Theorem 1.4, restated)**.** *For any $n, s \in \mathbb{N}$ and any $\epsilon > 0$, there is an ϵ-PRG for size-s De Morgan formulas with output length n that is computable in time $\mathrm{poly}(n)$ with seed length*

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

5.1 Shrinkage in expectation under truly random restrictions

Like the original PRG by Impagliazzo, Meka, and Zuckerman [IMZ19], the proof of Theorem 5.1 relies on the phenomenon that De Morgan formulas *shrink* under a random restriction. Let $\mathrm{L}(f)$ denote the size of the smallest De Morgan formula computing f (L for *leaf complexity*). Improving on work by Håstad [Hås98], Tal showed that formulas shrink in expectation under a truly random restriction.

Theorem 5.2 ([Tal14])**.** *Let f be a De Morgan formula of size s, and let $\boldsymbol{\rho}$ be a truly random p-regular restriction. Then $\mathbb{E}[\mathrm{L}(f\restriction_{\boldsymbol{\rho}})] \leq O\left(p^2 s + p\sqrt{s}\right) \leq O(p^2 s + 1)$.*

5.2 High-probability shrinkage of bounded-read De Morgan formulas

For our PRG, we would like to show some form of simplification *with high probability* under a *pseudorandom* restriction. Let us begin with the special case of read-t De Morgan formulas for small t, i.e., formulas where each variable appears in at most t leaves. Impagliazzo, Meka, and Zuckerman showed [IMZ19, Lemma 4.7] that indeed, under a pseudorandom restriction, such formulas shrink with high probability (see also [KRT17, Lemma 4.4]). Their bound is sufficient for our application (Theorem 5.1), but we take this opportunity to re-prove the result with improved parameters. We begin by recalling a structural lemma for De Morgan formulas that Tal proved [Tal14] building on the work by Impagliazzo, Meka, and Zuckerman [IMZ19].

Lemma 5.3 ([Tal14])**.** *Let f be a De Morgan formula of size $s \geq \ell$. There exist a read-once De Morgan formula F and De Morgan formulas $g_1, \ldots, g_m$ of size at most ℓ, with $m = O(s/\ell)$, such that for all x,*

$$f(x) = F(g_1(x), \ldots, g_m(x)).$$

Furthermore, if a variable x_i appears t times in f, then it appears at most $2t$ times in total among $g_1, \ldots, g_m$.

(The "furthermore" clause was not explicitly stated in Tal's work [Tal14], but it follows from an inspection of the construction.)

Lemma 5.4 (High-probability shrinkage of bounded-read De Morgan formulas)**.** *Let $p, \epsilon \in (0,1)$. Let f be a read-t De Morgan formula of size s, where $p^2 s \geq 1$. Let $\boldsymbol{\rho} \in \{0,1,\star\}^n$ be a k-wise independent p-regular restriction, where $k = p^{-2}\log(s/\epsilon)$. Then*

$$\Pr[\mathrm{L}(f\restriction_{\boldsymbol{\rho}}) \leq O(p^2 s + t \cdot p^{-4} \cdot \log(s/\epsilon))] \geq 1 - \epsilon.$$

Proof. Let $f(x) = F(g_1(x), \ldots, g_m(x))$ be the decomposition from Lemma 5.3 with $m \leq O(p^2 s)$ and $\mathrm{L}(g_i) \leq \ell = p^{-2}$. Since $k \geq \ell$, Theorem 5.2 implies that $\mathbb{E}[\mathrm{L}(g_i\restriction_{\boldsymbol{\rho}})] \leq s_0$ for some $s_0 = O(1)$. Now form a graph with vertices $g_1, \ldots, g_m$, where g_i is adjacent to g_j if there

is some variable that is read by both g_i and g_j. Since f is read-t, each variable appears at most $2t$ times in total among $g_1, \dots, g_m$, so the graph has maximum degree less than $2t\ell$. Therefore, there is a proper $(2t\ell)$-coloring of the graph.

Consider a single color class $S \subseteq [m]$ under such a coloring. Since S forms an independent set, the formulas g_i for $i \in S$ read disjoint sets of variables, and each g_i has at most ℓ leaves. Therefore, if we let $\mathbf{y}_i = \mathrm{L}(g_i\restriction_\rho)/\ell \in [0,1]$, then $\mathbf{y}_i$ for $i \in S$ are k'-wise independent for $k' = k/\ell = \log(s/\epsilon)$. By Theorem 4.5 with $\Delta = |S|s_0/\ell + 300\log(s/\epsilon)$, we have

$$\Pr\left[\sum_{i\in S} \mathrm{L}(g_i\restriction_\rho) > 2|S|s_0 + 300\ell\log(s/\epsilon)\right] \le \frac{\epsilon}{s}.$$

The number of colors is $2t\ell$ (which is at most s), so by the union bound,

$$\Pr\left[\sum_{i=1}^{m} \mathrm{L}(g_i\restriction_\rho) > 2ms_0 + 600t\ell^2\log(s/\epsilon)\right] \le \frac{\epsilon}{s}\cdot 2t\ell \le \epsilon.$$

Assuming that event does not occur,

$$\mathrm{L}(f\restriction_\rho) \le 2ms_0 + 600t\ell^2\log(s/\epsilon) = O(p^2 s + t\ell^2\log(s/\epsilon)). \quad \blacksquare$$

5.3 High-probability simplification of unbounded-read formulas

Let us turn to the general case of unbounded-read De Morgan formulas. Impagliazzo, Meka, and Zuckerman [IMZ19, Lemma 4.8] showed that with probability $1-\epsilon$ under a pseudorandom restriction, a size-s De Morgan formula shrinks to size

$$p^2 s \cdot \exp\left(O\left(\frac{\log(1/\epsilon)}{\log^{1/3} s}\right)\right).$$

This is meaningful in the moderate-error regime that they focused on (e.g., $\epsilon = 1/\mathrm{poly}(s)$), but the bound becomes trivial for smaller ϵ such as $\epsilon = 1/\mathrm{quasipoly}(s)$. Recall that we are aiming to design a PRG for errors as small as $2^{-s^{\Omega(1)}}$.

Unfortunately, as discussed in Section 2.1, the poor dependence on ϵ in Impagliazzo, Meka, and Zuckerman's result [IMZ19] is not an artifact of their analysis. Even under a *truly* random restriction, De Morgan formulas genuinely *do not shrink* with sufficiently high probability. To evade this obstacle, we are forced to consider a relaxed notion of "simplification." In particular, we study a hybrid decision tree model, defined next.

Definition 5.5. *A (D,s)-*`DMF-DT` *is a partial decision tree T of depth D where each internal node is labeled with a variable x_i and each leaf is labeled with either a size-s De Morgan formula or else $\perp$. The tree computes a function $T\colon \{0,1\}^n \to \{0,1,\perp\}$ in the natural way: follow the path in the partial decision tree until reaching a leaf defined by the input x. If the leaf is marked with $\perp$, output $\perp$, and otherwise output the value of the relevant formula evaluated on x. We define*

$$\mathsf{Err}(T) = \Pr[T(\mathbf{u}_n) = \perp].$$

Loosely speaking, we will now show that with probability $1-\epsilon$ under a pseudorandom restriction with $\star$-probability p, a size-s De Morgan formula simplifies to a `DMF-DT`, where the formulas at the leaves have size only $O(p^2 s)$ and the depth D of the tree (i.e., the maximum number of queries) is $O(p^{-6}\log(s/\epsilon))$. Intuitively, the hybrid decision tree model is helpful because we can query all the "heavy" variables, leaving us with a bounded-read formula.

In fact, since we will need to apply this lemma iteratively to further simplify the simpler function, we will prove more generally that `DMF-DT`s with size-s formulas at the leaves simplify

to DMF-DTs with size-$O(p^2 s)$-formulas at the leaves. Another technicality is that we merely achieve failure probability ϵ with respect to the pseudorandom choice of restriction *and* a truly random input to the restricted function. The precise statement follows.

Lemma 5.6 (High-probability simplification of DMF-DTs)**.** *Let T be a (D,s)-DMF-DT and let $p,\epsilon \in (0,1)$ with $p^2 s \geq 1$. For each restriction $\rho \in \{0,1,\star\}^n$, there is a (D',s')-DMF-DT T_ρ such that for all x, $T_\rho(x) \in \{(T\restriction_\rho)(x), \perp\}$, and*

$$s' \leq O(p^2 s) \qquad\qquad D' \leq D + O(p^{-6}\log(s/\epsilon)).$$

Furthermore, suppose $\boldsymbol{\rho}$ is sampled from a k-wise independent p-regular distribution, where $k \geq D + cp^{-6}\log(s/\epsilon)$ for a suitable constant c. Then

$$\mathop{\mathbb{E}}_{\boldsymbol{\rho}}[\mathsf{Err}(T_{\boldsymbol{\rho}})] \leq \mathsf{Err}(T) + \epsilon.$$

Proof. Let $x \in \{0,1\}^n$ be an input to T_ρ, and let $y = x \circ \rho$. The tree T_ρ begins by simulating the decision tree portion of $T(y)$, arriving at some leaf ℓ. If ℓ is labeled $\perp$, then the corresponding node of T_ρ is also a leaf labeled $\perp$. Otherwise, ℓ is labeled with a size-s De Morgan formula f. Let $t = \frac{p^6 \cdot s}{\log(s/\epsilon)}$. The tree T_ρ obtains the values of all variables y_i that appear more than t times in f (by querying x_i if $\rho_i = \star$), arriving at a leaf of T_ρ. Let τ denote the restriction describing these values of y, so $\tau_i \in \{y_i, \star\}$ and $f\restriction_\tau$ is a size-s read-t De Morgan formula. For these values of s, t, and ϵ, the bound in the conclusion of Lemma 5.4 is $Cp^2 s$ for some constant C. If $\mathrm{L}((f|_\tau)|_\rho) > Cp^2 s$, then the leaf of T_ρ is labeled $\perp$; otherwise, it is labeled with $(f\restriction_\tau)\restriction_\rho$.

By construction, $T_\rho(x) \in \{T(y), \perp\} = \{(T\restriction_\rho)(x), \perp\}$ and we can take $s' = Cp^2 s$. The number of variables that appear in f more than t times is at most s/t, so we can take $D' = D + s/t = D + O(p^{-6}\log(s/\epsilon))$. Finally, let us bound $\mathbb{E}[\mathsf{Err}(T_{\boldsymbol{\rho}})] = \Pr_{\boldsymbol{\rho},\mathbf{x}}[T_{\boldsymbol{\rho}}(\mathbf{x}) = \perp]$. We can determine which leaf of T is reached by $T_{\boldsymbol{\rho}}(\mathbf{x})$ by observing at most D coordinates of $\boldsymbol{\rho}$ and $\mathbf{x}$. The tree $T_{\boldsymbol{\rho}}$ will output $\perp$ if the leaf is labeled $\perp$ or if it is labeled f and $f\restriction_\tau$ fails to shrink. Since $k \geq D$, the probability that the leaf is labeled $\perp$ is precisely $\mathsf{Err}(T)$. Otherwise, we can determine the formula $f\restriction_\tau$ by observing at most s/t additional coordinates of $\boldsymbol{\rho}$ and $\mathbf{x}$. Conditioned on any values for those $D + s/t$ coordinates, the remaining coordinates of $\boldsymbol{\rho}$ are still k'-wise independent where $k' = k - (D + s/t) \geq p^{-2}\log(s/\epsilon)$. By Lemma 5.4, with respect to those remaining coordinates, the probability that $f\restriction_\tau$ fails to shrink is at most ϵ. Therefore, $\mathbb{E}[\mathsf{Err}(T_{\boldsymbol{\rho}})] \leq \mathsf{Err}(T) + \epsilon$. ∎

Ultimately we would like to use a restriction with $p \approx s^{-1/3}$, but with such a small p, the depth D' in Lemma 5.6 becomes completely trivial, and besides, the seed length of the k-wise independent restriction in Lemma 5.6 becomes too large. Following Impagliazzo, Meka, and Zuckerman [IMZ19], we circumvent this issue by composing several pseudorandom restrictions with a much larger value of p. Since we will not need to apply this next theorem iteratively, we assume for simplicity that we start with a De Morgan formula rather than a DMF-DT.

Theorem 5.7 (High-probability simplification of De Morgan formulas for small p)**.** *Let $n, s \in \mathbb{N}$, and let $p \in (0, 1/2)$ and $\epsilon > 0$. Let $\alpha = 1/\sqrt{\log s} = o(1)$. Then, there is a distribution over $\boldsymbol{\rho} \in \{0,1,\star\}^n$ that can be sampled in $\mathrm{poly}(n, s, \log(1/p))$ time using $O(s^\alpha \cdot \log^2(1/p) \cdot \log(1/\epsilon) \cdot \log n)$ truly random bits with the following properties.*

1. *$\boldsymbol{\rho}$ is p'-regular and k-wise independent where $p' \geq p$ and $k = s^\alpha \cdot \log(1/p) \cdot \log(1/\epsilon)$.*
2. *For any size-s De Morgan formula f, there is a (D,s')-DMF-DT $\mathbf{T}$ (jointly distributed with $\boldsymbol{\rho}$) such that*

$$\forall x, \Pr[\mathbf{T}(x) \in \{f\restriction_{\boldsymbol{\rho}}(x), \perp\}] = 1 \qquad s' \leq p^{2-O(\alpha)} \cdot s$$
$$\mathbb{E}[\mathsf{Err}(\mathbf{T})] \leq \epsilon \qquad D \leq s^\alpha \cdot \log(1/p) \cdot \log(1/\epsilon).$$

Proof. If $p < 2^{-s}$, the statement is trivial because the tree can simply query every variable that f reads, so assume $p \geq 2^{-s}$. Let q be the smallest power of two that is at least $s^{-\alpha/7}$, so $q = 2^{-\Theta(\sqrt{\log s})}$. Let $r = \lfloor \frac{\log(1/p)}{\log(1/q)} \rfloor$, so $q^r \in [p, p/q)$.

Sample r independent and identically distributed restrictions $\boldsymbol{\rho}_1, \ldots, \boldsymbol{\rho}_r \in \{0,1,\star\}^n$, where $\boldsymbol{\rho}_i$ is q-regular and k-wise independent for a value of k to be specified later. Set $\boldsymbol{\rho} = \boldsymbol{\rho}_r \circ \cdots \circ \boldsymbol{\rho}_1$, so indeed, $\boldsymbol{\rho}$ is p'-regular where $p' = q^r \geq p$. To define $\mathbf{T}$, think of f as a $(0,s)$-DMF-DT, and using the notation of Lemma 5.6, define

$$\mathbf{T} = (\cdots((f_{\boldsymbol{\rho}_1})_{\boldsymbol{\rho}_2})_{\boldsymbol{\rho}_3}\cdots)_{\boldsymbol{\rho}_r}.$$

In each iteration, we use failure probability ϵ/r.

Clearly, $\mathbf{T}(x) \in \{f\restriction_{\boldsymbol{\rho}}(x), \perp\}$ and $\mathbb{E}[\mathrm{Err}(\mathbf{T})] \leq \epsilon$. Furthermore, $\mathbf{T}$ is a (D, s')-DMF-DT, where

$$\begin{aligned} D &\leq O(rq^{-6}\log(rs/\epsilon)) \\ &\leq O(s^{6\alpha/7}\log(1/\epsilon)\log(1/p)\sqrt{\log s}) \\ &\leq O(s^{\alpha}\log(1/\epsilon)\log(1/p)) \end{aligned}$$

and

$$s' \leq O(q^2)^r \cdot s \leq 2^{O(r)}q^{2r}\cdot s \leq 2^{O(r)}\cdot p^2 \cdot q^{-2}\cdot s \leq p^{2-O(\frac{1}{\alpha \log s})}\cdot s = p^{2-O(\alpha)}\cdot s.$$

For each individual restriction, it suffices to take $k = O(D)$. By Claim 4.8, each $\boldsymbol{\rho}_i$ can be sampled using $O(k\log(n/q))$ truly random bits. Therefore, the total seed length is

$$O(rk\log(n/q)) \leq O(s^{\alpha}\log(1/\epsilon)\log^2(1/p)\log n)\,. \quad \blacksquare$$

5.4 The PRG construction

Now we are ready to construct the PRG of Theorem 5.1. We follow the framework of Impagliazzo, Meka, and Zuckerman [IMZ19]. Sample independent copies $\boldsymbol{\rho}_1, \ldots, \boldsymbol{\rho}_t$ of the restriction from Theorem 5.7, with

$$p = s^{-1/3}\cdot \log^{1/3}(1/\epsilon), \tag{3}$$

error parameter ϵ/t, and $t = O(p^{-1}\log n + \log(1/\epsilon))$. Let s' and D be the values from Theorem 5.7, and let $G\colon \{0,1\}^d \to \{0,1\}^n$ be a $(D+s')$-wise independent generator. Sample a $(2400p^{-1})$-wise independent string $\mathbf{w}$. Let $\mathsf{Ext}\colon \{0,1\}^m \times \{0,1\}^{d_{\mathsf{Ext}}} \to \{0,1\}^d$ be a $(k_{\mathsf{Ext}}, \epsilon_{\mathsf{Ext}})$-extractor, where k_{Ext}, ϵ_{Ext} and m will be specified later. Sample independent uniform random strings $\mathbf{a} \in \{0,1\}^m$ and $\mathbf{y}_1, \ldots, \mathbf{y}_t \in \{0,1\}^{d_{\mathsf{Ext}}}$. Let $\mathbf{z}_i = \mathsf{Ext}(\mathbf{a}, \mathbf{y}_i)$. Our PRG outputs

$$\mathbf{w} \oplus \bigoplus_{i=1}^{t}(G(\mathbf{z}_i)\circ \boldsymbol{\rho}_i),$$

i.e., we use $G(\mathbf{z}_i)$ to fill in the $\star$ coordinates of $\boldsymbol{\rho}_i$ and then take the bitwise XOR of the resulting t strings and $\mathbf{w}$.

The proof of correctness is a hybrid argument. Sample independent uniform strings $\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(t)}$, and define hybrid distributions $\mathbf{h}_0, \ldots, \mathbf{h}_t$ by

$$\mathbf{h}_j = \mathbf{w} \oplus \left(\bigoplus_{i=1}^{j}(\mathbf{u}^{(i)}\circ\boldsymbol{\rho}_i)\right) \oplus \left(\bigoplus_{i=j+1}^{t} G(\mathbf{z}_i)\circ\boldsymbol{\rho}_i\right).$$

Claim 5.8. *Let f be a size-s De Morgan formula, and let $j \in [t]$. Then*

$$|\,\mathbb{E}[f(\mathbf{h}_{j-1})] - \mathbb{E}[f(\mathbf{h}_j)]| \leq \tfrac{\epsilon}{t} + 3\cdot 2^D \cdot \epsilon_{\mathsf{Ext}}.$$

Proof. Define the random variable $\mathbf{e} = \mathbf{w} \oplus \left(\bigoplus_{i=1}^{j-1}(\mathbf{u}^{(i)} \circ \boldsymbol{\rho}_i)\right) \oplus \left(\bigoplus_{i=j+1}^{t} G(\mathbf{z}_i) \circ \boldsymbol{\rho}_i\right)$ which conveniently gives us

$$\mathbf{h}_{j-1} = \mathbf{e} \oplus (G(\mathbf{z}_j) \circ \boldsymbol{\rho}_j)$$
$$\mathbf{h}_j = \mathbf{e} \oplus (\mathbf{u}^{(j)} \circ \boldsymbol{\rho}_j).$$

Note that $\mathbf{e}$ and $\mathbf{a}$ are *not* independent, but $\mathbf{e}$ is nevertheless independent of $\boldsymbol{\rho}_j$ and $\mathbf{y}_j$. Let $f^{\oplus \mathbf{e}}(x) = f(\mathbf{e} \oplus x)$, so $f^{\oplus \mathbf{e}}$ can be computed by a size-s De Morgan formula. Let $\mathbf{T}$ be the DMF-DT from Theorem 5.7 associated with $\boldsymbol{\rho}_j$ and $f^{\oplus \mathbf{e}}$, so $\mathbf{T}(x) \in \{(f^{\oplus \mathbf{e}})|_{\boldsymbol{\rho}_j}(x), \perp\}$. Define $\mathbf{r}\colon \{0,1\}^n \to [2^D]$ by letting $\mathbf{r}(x)$ be the index of the leaf reached by $\mathbf{T}(x)$. For each $\ell \in [2^D]$, define

$$\mathbf{g}_\ell^-(x) = 1 \iff (\mathbf{r}(x) = \ell) \wedge \mathbf{T}(x) = 1$$
$$\mathbf{g}_\ell^+(x) = 1 \iff (\mathbf{r}(x) = \ell) \wedge \mathbf{T}(x) \in \{1, \perp\}.$$

Now, $\mathbf{g}_\ell^+$ (as a function) is a random variable, since $\mathbf{g}_\ell^+$ depends on $\mathbf{e}$ and $\boldsymbol{\rho}_j$. Let us bound the support size of that random variable. For each fixed ℓ, to describe the function $\mathbf{g}_\ell^+$, we could specify the sequence of variables and values on the path from the root to leaf ℓ of $\mathbf{T}$, and then we could specify the label of that leaf (either $\perp$ or a size-s' De Morgan formula). In total, that requires at most $O(D \log n + s' \log n) = O(s' \log n)$ bits.[17]

Intuitively, this means that $\mathbf{a}$ has a lot of entropy even given $\mathbf{g}_\ell^+$. By the extractor guarantee, this should imply that $\mathbf{z}_j = \mathsf{Ext}(\mathbf{a}, \mathbf{y}_i)$ is near uniform even given $\mathbf{g}_\ell^+$, and hence $\mathbf{g}_\ell^+(G(\mathbf{z}_j)) \approx \mathbf{g}_\ell^+(G(\mathbf{u}_d))$.

The cleanest way to make this sort of argument rigorous is to use the notion of *conditional min-entropy* [DORS08]. Recall that for two random variables such as $\mathbf{a}$ and $\mathbf{g}_\ell^+$, the min-entropy of $\mathbf{a}$ given $\mathbf{g}_\ell^+$ is defined as

$$\widetilde{H}_\infty(\mathbf{a} \mid \mathbf{g}_\ell^+) = \log\left(\frac{1}{\mathbb{E}_{g \sim \mathbf{g}_\ell^+}[\max_a \Pr[\mathbf{a} = a \mid \mathbf{g}_\ell^+ = g]}\right).$$

Conditional min-entropy satisfies a *chain rule* [DORS08, Lemma 2.2], giving

$$\widetilde{H}_\infty(\mathbf{a} \mid \mathbf{g}_\ell^+) \geq H_\infty(\mathbf{a}) - \log(\mathrm{supp}(\mathbf{g}_\ell^+)) \geq m - O(s' \log n).$$

Furthermore, extractors can extract randomness from sources with high conditional min-entropy. In particular, if we set k_{Ext} to be that same quantity $m - O(s' \log n)$, then [Vad12, Problem 6.8]

$$(\mathbf{g}_\ell^+, \mathbf{z}_j) \sim_{3\epsilon_{\mathsf{Ext}}} (\mathbf{g}_\ell^+, \mathbf{u}),$$

where $\mathbf{u}$ is a uniform random d-bit string independent of $\mathbf{g}_\ell^+$ and $\sim_\gamma$ denotes γ-closeness in total variation distance. Deterministic processing can only decrease total variation distance, so

$$\left|\mathbb{E}[\mathbf{g}_\ell^+(G(\mathbf{z}_j))] - \mathbb{E}[\mathbf{g}_\ell^+(G(\mathbf{u}))]\right| \leq 3\epsilon_{\mathsf{Ext}}.$$

Furthermore, $\mathbf{g}_\ell^+$ only reads $D + s'$ variables, so G perfectly fools $\mathbf{g}_\ell^+$. Therefore,

$$\left|\mathbb{E}[\mathbf{g}_\ell^+(G(\mathbf{z}_j))] - \mathbb{E}[\mathbf{g}_\ell^+(\mathbf{u}^{(j)})]\right| \leq 3\epsilon_{\mathsf{Ext}}.$$

The same argument applies just as well to $\mathbf{g}_\ell^-$, showing that

$$\left|\mathbb{E}[\mathbf{g}_\ell^-(G(\mathbf{z}_j))] - \mathbb{E}[\mathbf{g}_\ell^-(\mathbf{u}^{(j)})]\right| \leq 3\epsilon_{\mathsf{Ext}}.$$

[17] We may assume without loss of generality that $s \leq n^3$, since otherwise the theorem is trivial.

Altogether, we have

$$\begin{aligned}\mathbb{E}[f^{\oplus \mathbf{e}}(G(\mathbf{z}_j) \circ \boldsymbol{\rho}_j)] &\leq \sum_{\ell=1}^{2^D} \mathbb{E}[\mathbf{g}_\ell^+(G(\mathbf{z}_j))] \\ &\leq \mathbb{E}\left[\sum_{\ell=1}^{2^D} \mathbf{g}_\ell^+(\mathbf{u}^{(j)})\right] + 2^D \cdot 3\epsilon_{\mathsf{Ext}} \\ &= \mathbb{E}\left[\sum_{\ell=1}^{2^D} \mathbf{g}_\ell^-(\mathbf{u}^{(j)})\right] + \mathbb{E}[\mathsf{Err}(\mathbf{T})] + 2^D \cdot 3\epsilon_{\mathsf{Ext}} \\ &\leq \mathbb{E}[f^{\oplus \mathbf{e}}(\mathbf{u}^{(j)} \circ \boldsymbol{\rho}_j)] + \frac{\epsilon}{t} + 2^D \cdot 3\epsilon_{\mathsf{Ext}}.\end{aligned}$$

A symmetric argument proves a bound in the other direction. Since $f^{\oplus \mathbf{e}}(G(\mathbf{z}_j) \circ \boldsymbol{\rho}_j) = f(\mathbf{h}_{j-1})$ and $f^{\oplus \mathbf{e}}(\mathbf{u}^{(j)} \circ \boldsymbol{\rho}_j) = f(\mathbf{h}_j)$, this completes the proof. ■

Claim 5.8 showed that the hybrid distributions are indistinguishable by f. Clearly, $\mathbf{h}_0$ is our pseudorandom distribution. Next, we show that $\mathbf{h}_t$ is statistically close to $\mathbf{u}_n$.

Claim 5.9. $\mathbf{h}_t \sim_\epsilon \mathbf{u}_n$.

Proof. Suppose we have already sampled $\boldsymbol{\rho}_1, \ldots, \boldsymbol{\rho}_{j-1}$. Let $\mathbf{I}$ be the set of coordinates in $[n]$ that are not yet covered by a $\star$, i.e.,

$$\mathbf{I} = [n] \setminus \left(\bigcup_{i=1}^{j-1} \boldsymbol{\rho}_i^{-1}(\star) \right).$$

Let us say that $\boldsymbol{\rho}_j$ is a *success* if either $|\mathbf{I}| \leq 2400p^{-1}$ or $|\boldsymbol{\rho}_j^{-1}(\star) \cap \mathbf{I}| \geq p|\mathbf{I}|/2$. Recall from Theorem 5.7 that $\boldsymbol{\rho}_j$ is 4-wise independent (indeed, k-wise independent for $k \gg 4$) and p'-regular for some $p' \geq p$. Therefore, by Theorem 4.5 with $\Delta = p'|I|/2$, conditioned on any values of $\boldsymbol{\rho}_1, \ldots, \boldsymbol{\rho}_{j-1}$, the probability that $\boldsymbol{\rho}_j$ is a success is at least $15/16$. This implies that with high probability, there will be at least $\Omega(t)$ successes. Indeed, by Azuma's inequality [DRV10, Lemma B.1], there will be at least $t/2$ successes, except with probability $e^{-\Omega(t)} \leq \epsilon$ (recalling that $t = O(p^{-1} \log n + \log(1/\epsilon))$).

Suppose now that there are at least $t/2$ successes. We start with n coordinates not covered by $\star$, and whenever there is a success, either we are down to $2400p^{-1}$ coordinates not covered, or else the number of coordinates not covered decreases by a factor of $(1 - p/2)$. If the latter case happens every time, then every coordinate is covered, because $(1 - p/2)^{t/2} \cdot n < 1$ (here we use $t \geq \Omega(p^{-1} \log n)$). Therefore, either way we must eventually cover all but $2400p^{-1}$ coordinates. In that case, with respect to the randomness of $\mathbf{w}$ and $\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(t)}$, $\mathbf{h}_t$ is clearly uniformly random. ■

Proof of Theorem 5.1. By Claim 5.8 and the triangle inequality,

$$|\mathbb{E}[f(\mathbf{h}_0)] - \mathbb{E}[f(\mathbf{h}_t)]| \leq \epsilon + 2^D \cdot 3t\epsilon_{\mathsf{Ext}}.$$

By Claim 5.9, this implies that

$$|\mathbb{E}[f(\mathbf{h}_0)] - \mathbb{E}[f]| \leq 2\epsilon + 2^D \cdot 3t\epsilon_{\mathsf{Ext}}.$$

Recall that $\mathbf{h}_0$ is the output distribution of our PRG. We shall set $\epsilon_{\mathsf{Ext}} = \epsilon 2^{-D}/t$, so the total error is $O(\epsilon)$.

Finally, let us bound the seed length of our PRG. We have $\log(1/p) < \log n$ and $\log(1/\delta) = O(\log(n/\epsilon))$. Therefore, each $\boldsymbol{\rho}_i$ costs fewer than

$$s^{\alpha} \cdot \log(1/\epsilon) \cdot \mathrm{polylog}(n)$$

truly random bits, where $\alpha = 1/\sqrt{\log s} = o(1)$. Furthermore, using Eq. (3),

$$s' = p^{2-O(\alpha)} s = s^{1/3+O(\alpha)} \log^{2/3}(1/\epsilon).$$

Therefore, the generator G has seed length

$$d = O((D+s')\log n) \leq s^{\alpha} \log(1/\epsilon) \cdot \mathrm{polylog}(n) + s^{1/3+O(\alpha)} \cdot \log^{2/3}(1/\epsilon).$$

We shall take Ext to be the Guruswami-Umans-Vadhan extractor [GUV09], so we can take $k_{\mathsf{Ext}} = O(d)$. As mentioned previously, we also need $k_{\mathsf{Ext}} = m - O(s' \log n)$, so we shall take $m = O(d)$. Of course, to sample $\mathbf{a}$ we need m truly random bits. The seed length d_{Ext} of this extractor is

$$O(\log(m/\epsilon_{\mathsf{Ext}})) = O(\log(m) + D + \log(t)) = s^{\alpha} \log(1/\epsilon) \cdot \mathrm{polylog}(n).$$

Finally, the $O(p^{-1})$-wise independent distribution $\mathbf{w}$ costs another $O(p^{-1} \log n)$ truly random bits. Summing up, the total seed length is

$$s^{1/3+O(\alpha)} \cdot \log^{2/3}(1/\epsilon) + (p^{-1} + \log(1/\epsilon)) \cdot s^{\alpha} \cdot \log(1/\epsilon) \cdot \mathrm{polylog}(n),$$

which is bounded by

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

6 A PRG for LTF circuits of super-linear size

In this section we prove Theorem 1.1. We first present, in Section 6.1, a pseudorandom restriction procedure that simplifies any LTF circuit into a suitable hybrid model, and fails with tiny probability. Then, in Section 6.2, we construct a low-error PRG for the latter hybrid model. And we combine these two parts into a PRG using the proof framework of Ajtai and Wigderson [AW85] in Section 6.3.

6.1 Low-error pseudorandom restrictions for LTF circuits

In this section we present our pseudorandom restriction procedure for LTF circuits. We begin, in Section 6.1.1, by defining the hybrid model to which any LTF circuit will be simplified with very high probability, which is a generalization of decision trees and that we call an LTF-DT. Then, in Section 6.1.2 we analyze the effect of a carefully structured restriction on a *single* LTF function (i.e., a single gate in the circuit). In Section 6.1.3 we show how to simplify any LTF circuit to an LTF-DT whose leaves are labeled by LTF circuits of smaller depth, with very high probability. Anticipating iterative applications of this procedure, in Section 6.1.4 we show how to simplify an LTF-DT whose leaves are labeled by LTF circuits into another LTF-DT whose leaves are labeled by LTF circuits of smaller depth. Finally, in Section 6.1.5 we show how to iteratively apply the foregoing procedure in order to simplify an LTF circuit into an LTF-DT whose leaves are labeled by LTF functions (i.e., of LTF circuits of depth one).

Actually, as we will explain below, we show that each LTF circuit can be *approximated* by an LTF-DT, and additionally that the approximation error of the LTF-DT can be decided by a relatively simple function (i.e., at each leaf of the LTF-DT the approximation error can be decided by the sum of polynomially-many LTFs). The fact that the approximation error can be decided in this way will be crucial to our PRG construction, as explained in Section 6.3.

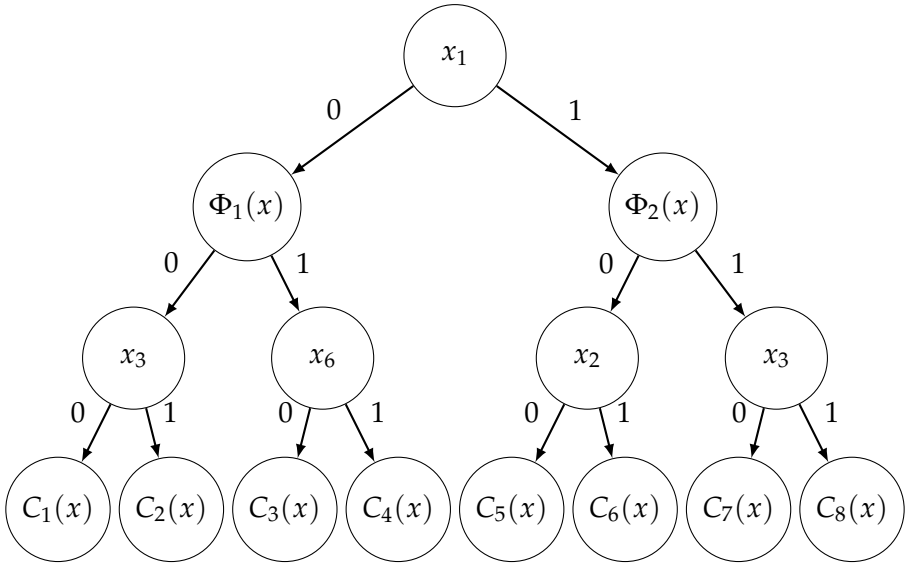

Figure 1: An $(n, 1, 2, d, w)$-LTF-DT. Each Φ_i is an LTF, and each C_i is an LTF circuit with depth d and at most w wires on variables $x_1, \ldots, x_n$. In the more refined hybrid model presented in Definition 6.2, each leaf ℓ is labeled not only with C_i but also with a set $\mathcal{E}_\ell$ of LTFs (i.e., with error indicators).

6.1.1 The hybrid computational model

Let us now formally define LTF decision trees. We first present a simplified definition and then expand it to the full definition.

Definition 6.1 (LTF decision tree, see Figure 1)**.** *For $n \in \mathbb{N}$, we say that a binary tree T is an (n, M, D, d, w)* LTF decision tree *(or (n, M, D, d, w)-LTF-DT, in short) if:*

1. *Each internal node in T is labeled either by a variable x_i (for some $i \in [n]$) or by a linear threshold function $\{0,1\}^n \to \{0,1\}$.*
2. *Each leaf ℓ of T is labeled by an LTF circuit C_ℓ over n variables of depth d and with at most w wires, and none of the variables queried on the path to ℓ appear as input variables in C_ℓ.*
3. *In the path to each leaf in the tree there are at most M nodes labeled with an LTF and at most D nodes labeled with a variable. (In particular, the depth of T is at most $D + M$.)*

An (n, M, D, d, w)-LTF-DT computes a function $T\colon \{0,1\}^n \to \{0,1\}$ in the natural way: Given input $x \in \{0,1\}^n$, we traverse from the root of T along the path that corresponds to the evaluation of the function in each node (i.e., either an LTF or a variable) until we reach a leaf $\ell = \ell(x)$ that is labeled by an LTF circuit C_ℓ, and we output $C_\ell(x)$.

As mentioned above, we will actually argue that each LTF circuit is *approximated* by an LTF-DT, and moreover that the approximation error can be tested (at each leaf of the LTF-DT) by a "simple function". This is formalized in the following definition:

Definition 6.2 (LTF decision tree with error indicators)**.** *We define a (n, M, D, d, w, e)* LTF decision tree with error indicators *(or (n, M, D, d, w, e)-LTF-DT, in short) to be an (n, M, D, d, w)-LTF-DT where each leaf ℓ of the tree is labeled with a set $\mathcal{E}_\ell$ of LTF functions with $|\mathcal{E}_\ell| \leq e$ (in addition to the LTF circuit C_ℓ that labels ℓ as before).*

The interpretation is, if some $\Phi \in \mathcal{E}_\ell$ outputs 1, that means something has gone wrong and the output of the tree is not reliable. An (n, M, D, d, w, e)-LTF-DT computes a function $T\colon \{0,1\}^n \to \{0,1\}$ as before: given input $x \in \{0,1\}^n$ we traverse the corresponding path in the tree until reaching a leaf $\ell = \ell(x)$, and then we output $C_\ell(x)$.

Definition 6.3 (error functions of an LTF-DT)**.** *Given an* LTF-DT $T\colon \{0,1\}^n \to \{0,1\}$*, we define its* error-indicator function *by* $\mathsf{Err}(T,x) = \sum_{\Phi\in\mathcal{E}_{\ell(x)}} \Phi(x)$*. We define the* error *of* T *by* $\mathsf{Err}(T) = \mathbb{E}_{x\sim\mathbf{u}_n}[\mathsf{Err}(T,x)]$.

Recall that when T was a DMF-DT, we defined $\mathsf{Err}(T)$ to be the *probability* of error on a random input. Now that T is an LTF-DT, we have defined $\mathsf{Err}(T)$ to be the *expected number* of errors on a random input.

Definition 6.4 (consistency of an LTF-DT with a function)**.** *Given a function* $f\colon \{0,1\}^n \to \{0,1\}$ *and an* LTF-DT $T\colon \{0,1\}^n \to \{0,1\}$*, we say that* T *is* consistent with f *if for every* $x \in \{0,1\}^n$*,*

$$T(x) \neq f(x) \implies \mathsf{Err}(T,x) \geq 1\,.$$

6.1.2 Simplification of a single LTF under a random restriction

To prove our result we will rely on a result asserting that any *single* LTF becomes close a constant, with decent probability, under a restriction that randomly assigns values to a fixed "good" set of variables. A similar result was proved by Chen, Santhanam, and Srinivasan [CSS18, Lem. 4.4], who analyzed the case of a truly random restriction, and a derandomized and refined version was proved in [Tel18, Prop. 3.8].

To get some intuition for the result, consider the illustrative example of the simple majority function on n bits and of a random restriction. With extremely high probability, the ℓ_2 norm of the living variables is $\approx \sqrt{p\cdot n}$; and by a suitable anti-concentration result (i.e., the Berry-Esseen Theorem), the probability that a random fixing of the rest of the variables has absolute value less than $t\cdot\sqrt{p\cdot n}$ is at most $O(t\cdot\sqrt{p})$. Thus, with probability at least $1-O(t\cdot\sqrt{p})$ the restricted function has variables with ℓ_2 norm $\approx\sqrt{p\cdot n}$ and "threshold" more than $t\cdot\sqrt{p}$, in which case it is $\exp(-t)$-close to a constant function (see [Tel18, Fact 5.3] for details).

Indeed, the result of [CSS18] asserts that the same phenomenon happens for *every* LTF rather than just for the majority function. We will need a variant of the derandomized version of [Tel18] for this result: For any LTF, we will condition on a corresponding "good" choice of variables to keep alive (where the meaning of "good" appears below), and assert that with decent probability under a uniform choice of values for fixed variables, the LTF becomes close to a constant.[18] (We will also parametrize the result more carefully than in [Tel18].)

The proof below of this result closely follows [CSS18; Tel18]. In high-level, following an idea of [Ser07], we partition the variables of the LTF into the "head" and the "tail," where the head consists of the variables with the largest weights in absolute value. (The precise partition to head and tail relies on the critical index of the LTF, as defined in Definition 4.10.) Intuitively, for any fixing of the head variables, the residual function on the tail variables behaves approximately like a majority function, whose behavior under a random restriction was analyzed above. In addition, the weights *inside the head* behave in a structured way (i.e., they essentially decrease exponentially), which makes them amenable to analysis. The proof thus proceeds by a case analysis, depending on the relative total weight of the head.

Proposition 6.5 (random restriction lemma for a single LTF and a fixed set of living variables)**.** *Let* $n\in\mathbb{N}$*, let* $\mu,\mu',\epsilon>0$ *such that* $\mu \leq \frac{1}{4\sqrt{2\log(2/\epsilon)}}$*, and let* $\lambda = \frac{10\log^2(1/\mu)}{\mu^2}$*. Let* $\Phi\colon\{0,1\}^n\to\{0,1\}$ *be an* LTF *with* μ*-critical index* h*, and assume that the weights vector* $w\in\mathbb{R}^n$ *of* $\Phi=(w,\theta)$ *satisfies* $|w_1|\geq|w_2|\geq\ldots\geq|w_n|$*. Let* $I\subseteq[n]$ *such that:*

1. *If* $h>\lambda$ *we require that* $I\cap[\lambda]=\emptyset$*.*

[18] The original statement in [Tel18] also referred to a pseudorandom choice of values for the fixed variables, but we only need a statement that refers to a pseudorandom choice of variables to fix and to a uniform choice of values for the fixed variables.

2. *If $h \leq \lambda$ we require that $I \cap [h] = \emptyset$ and $\|w_I\|_2 \leq \mu' \cdot \left\|w_{[n]\setminus[h]}\right\|_2$.*

Then, the probability over a uniform choice of $\mathbf{z} \in \{0,1\}^{[n]\setminus I}$ that the restricted function $\Phi\restriction_{I,\mathbf{z}}$ is ϵ-close to a constant is at least $1 - O(\mu + \mu' \cdot \sqrt{\log(1/\epsilon)})$.

Proof. Let $J = [n] \setminus I$ be the set of variables fixed by $\mathbf{z}$, and let $T = [n] \setminus [\min\{h, \lambda\}]$ be the "tail" of variables following the "head" of the first $\min\{h, \lambda\}$ variables. For convenience, for each string $x \in \{0,1\}^*$, let $e[x] \in \{-1,1\}^{|x|}$ be the string $e[x]_i = (-1)^{x_i}$. Letting $w' = -\frac{1}{2}w$, there is some θ' such that

$$\Phi(x) = 1 \iff \langle w', e[x] \rangle > \theta'.$$

Note that the μ-critical index of w' is h as well. We will prove that with probability at least $1 - O(\mu + \mu' \cdot \log(1/\epsilon))$ over choice of $\mathbf{z} \in \{0,1\}^J$,

$$\left\langle w'_J, e[\mathbf{z}] \right\rangle \notin \theta' \pm \sqrt{2\log(2/\epsilon)} \cdot \left\|w'_I\right\|_2 .$$

Using Hoeffding's inequality (Theorem 4.3), whenever this happens the restricted function $\Phi\restriction_{I,\mathbf{z}}$ is ϵ-close to a constant. Our proof of this claim relies on a case analysis.

The case of $h > \lambda$. For this case we use the following claim (the original notation is modified to be consistent with our notation):

Claim 6.5.1 (see [Tel18, Claim 5.7.1 in full version], following [CSS18] and [DGJSV10, Lemma 5.5]). *For any $r \in \mathbb{N}$, let $\lambda_{r,\mu} = \frac{4r \cdot \ln(3/\mu^2)}{\mu^2}$. Then, for any $J \supseteq [\lambda_{r,\mu}]$, the probability under uniform choice of $\mathbf{z} \in \{0,1\}^J$ that $\left\langle w'_J, e[\mathbf{z}] \right\rangle \in \theta' \pm \frac{1}{4\mu} \cdot \left\|w'_{>\lambda_{r,\mu}}\right\|_2$ is at most 2^{-r}.*

We use the claim above with parameter $r = \log(1/\mu)$, relying on our hypothesis that $\lambda = 10 \cdot \frac{\log^2(1/\mu)}{\mu^2} > \lambda_{r,\mu}$, to deduce that $\Pr_{\mathbf{z}}\left[\left\langle w'_J, e[\mathbf{z}] \right\rangle \notin \theta' \pm (1/4\mu) \cdot \left\|w'_{>\lambda_{r,\mu}}\right\|_2\right] \geq 1 - \mu$. Now, observe that

$$\begin{aligned}
\frac{1}{4\mu} \cdot \left\|w'_{>\lambda_{r,\mu}}\right\|_2 &\geq \sqrt{2\log(2/\epsilon)} \cdot \left\|w'_{>\lambda_{r,\mu}}\right\|_2 && (\mu \leq \tfrac{1}{4\sqrt{2\log(2/\epsilon)}}) \\
&> \sqrt{2\log(2/\epsilon)} \cdot \left\|w'_T\right\|_2 && (\lambda > \lambda_{r,\mu}) \\
&\geq \sqrt{2\log(2/\epsilon)} \cdot \left\|w'_I\right\|_2 , && (I \subseteq T)
\end{aligned}$$

and hence the probability over $\mathbf{z}$ that $\left\langle w'_J, e[\mathbf{z}] \right\rangle \notin \theta' \pm \sqrt{2\log(2/\epsilon)} \cdot \|w'_I\|_2$ is at least $1 - \mu$.

The case of $h \leq \lambda$. For this case, observe that for any fixed value for $\mathbf{z}_{[h]}$, the event $\left\langle w'_J, e[\mathbf{z}] \right\rangle \in \theta' \pm \sqrt{2\log(2/\epsilon)} \cdot \|w'_I\|_2$ happens if and only if $\left\langle w'_{T\setminus I}, e[\mathbf{z}_{T\setminus I}] \right\rangle \in \theta'' \pm \sqrt{2\log(2/\epsilon)} \cdot \|w'_I\|_2$, where $\theta'' = \theta' - \left\langle w'_{[h]}, e[\mathbf{z}_{[h]}] \right\rangle$. Relying on the hypothesis that $\|w_I\|_2 \leq \mu' \cdot \|w_T\|_2$ (hence $\|w'_I\|_2 \leq \mu' \cdot \|w'_T\|_2$), to upper-bound the probability of the latter event it suffices to upper-bound the probability of the event $\left\langle w'_{T\setminus I}, e[\mathbf{z}_{T\setminus I}] \right\rangle \in \theta'' \pm \sqrt{2\log(2/\epsilon)} \cdot \mu' \cdot \|w'_T\|_2$. We do so using the following claim, whose proof amounts to an application of the Berry-Esseen theorem:

Claim 6.5.2 (see [Tel18, Lemma 5.5 in full version]). *Assume that w'_T is μ-regular and that $\mu' \leq 3/4$, and let $I \subseteq T$ such that $\|w'_I\|_2 < \mu' \cdot \|w'_T\|_2$. Then, for any $\theta'' \in \mathbb{R}$ and $r > 0$, the probability over uniform choice of $\mathbf{z}' \in \{0,1\}^{T\setminus I}$ that $\left\langle w'_{T\setminus I}, e[\mathbf{z}'] \right\rangle \in \theta'' \pm r \cdot \mu' \cdot \|w'_T\|_2$ is at most $O(r \cdot \mu' + \mu)$.*

We invoke the claim above with $r = \sqrt{2\log(2/\epsilon)}$, relying on the fact that w'_T is μ-regular and that $\mu' \leq 3/4$.[19] We deduce that the probability that $\left\langle w'_{T\setminus I}, e[\mathbf{z}'] \right\rangle \in \theta'' \pm \sqrt{2\log(2/\epsilon)} \cdot \mu' \cdot \|w'_T\|_2$ is $O(\mu + \mu' \cdot \sqrt{\log(1/\epsilon)})$. By the hypothesis that $\|w'_I\|_2 \leq \mu' \cdot \|w'_T\|_2$, the probability that $\left\langle w'_{T\setminus I}, e[\mathbf{z}'] \right\rangle \in \theta'' \pm \sqrt{2\log(2/\epsilon)} \cdot \|w'_I\|_2$ is $O(\mu + \mu' \cdot \sqrt{\log(1/\epsilon)})$. Since this holds for any fixed value for $\mathbf{z}_{[h]}$, it follows that the probability over $\mathbf{z} \in \{0,1\}^J$ that $\left\langle w'_J, e[\mathbf{z}] \right\rangle \in \theta' \pm \sqrt{2\log(2/\epsilon)} \cdot \|w'_I\|_2$ is at most $O(\mu + \mu' \cdot \sqrt{\log(1/\epsilon)})$. ■

The next proposition considers a restriction where we assign truly random values to a pseudorandom subset of the variables, as well as to the variables with the largest weights (in absolute value). Under such a restriction, the proposition asserts that the LTF is likely to become extremely biased. Furthermore, the bad event (i.e., the event that the LTF does not become extremely biased) is broken up into two bad events: a bad set of living variables, or a bad assignment. The probability of a bad assignment is moderately low, and crucially, the probability of a bad set of living variables is extremely low.

Proposition 6.6 (pseudorandom restriction lemma for a single LTF)**.** *For any $\epsilon, \gamma > 0$, there is a value*

$$\lambda = O(\gamma^{-2} \cdot \log^2(1/\epsilon) \cdot \log^2(\log(1/\epsilon)/\gamma)) = \widetilde{O}(\gamma^{-2} \cdot \log^2(1/\epsilon)) \tag{4}$$

such that the following holds. Let $\Phi\colon \{0,1\}^n \to \{0,1\}$ be an LTF with weights vector $w \in \mathbb{R}^n$, and assume $|w_1| \geq |w_2| \geq \cdots \geq |w_n|$. Let $A \subseteq [n] \setminus [\lambda]$ and let $\mathbf{I}$ be a p-regular $\log(1/\epsilon)$-wise independent subset of A. Sample $\mathbf{z} \in \{0,1\}^{[n]\setminus\mathbf{I}}$ uniformly at random. Then with probability $1-\epsilon$ over $\mathbf{I}$,

$$\Pr_{\mathbf{z}}[\Phi\restriction_{\mathbf{I},\mathbf{z}} \text{ is not } \epsilon\text{-close to a constant}] \leq \gamma + O\left(\sqrt{p\log(1/\epsilon)}\right).$$

Proof. We wish to apply Proposition 6.5 with

$$\mu = \Theta\left(\frac{\gamma}{\log(1/\epsilon)}\right).$$

For this value of μ, the parameter λ in Proposition 6.5 is indeed given by Equation 4. We must verify that with high probability $\mathbf{I}$ satisfies the Proposition's hypotheses. Let h be the μ-critical index of Φ. When $h > \lambda$, $\mathbf{I}$ is guaranteed to satisfy the hypotheses, because $\mathbf{I} \cap [\lambda] = \emptyset$ by design. Assume now that $h \leq \lambda$. Let $T = [n] \setminus [h]$, so $\mathbf{I} \subseteq A \subseteq T$. We need to show that with high probability over $\mathbf{I}$, we have $\|w_{\mathbf{I}}\| \leq \mu' \|w_T\|$ for some small parameter μ'. Indeed, we claim that

$$\Pr_{\mathbf{I}}\left[\|w_{\mathbf{I}}\|_2 \leq O\left(\sqrt{p} + \mu\sqrt{\log(1/\epsilon)}\right) \cdot \|w_T\|_2\right] \geq 1-\epsilon. \tag{5}$$

To prove it, for each $i \in A$, let $\mathbf{a}_i$ be the random variable

$$\mathbf{a}_i = \frac{w_i^2}{\mu^2\|w_T\|_2^2} \cdot \mathbb{1}_{i\in\mathbf{I}}.$$

Note that, since $I \subseteq A$, we have that $\mu^2 \cdot \sum_{i\in A} \mathbf{a}_i = \frac{\|w_{\mathbf{I}}\|_2^2}{\|w_T\|_2^2}$. Since w_T is μ-regular, $\mathbf{a}_i \in [0,1]$. Furthermore,

$$\mathbb{E}\left[\sum_{i\in A} \mathbf{a}_i\right] = \frac{p\|w_A\|_2^2}{\mu^2\|w_T\|_2^2} \leq \frac{p}{\mu^2}.$$

[19] Note that we can assume $\mu' \leq 3/4$ without loss of generality, otherwise the statement that we are trying to prove is trivial (because our claimed error bound $O(\mu' \cdot \sqrt{\log(1/\epsilon)} + \mu)$ is larger than one).

By Theorem 4.5 with $\Delta = p/\mu^2 + 300\log(1/\epsilon)$,

$$\Pr\left[\sum_{i\in A} \mathbf{a}_i \geq 2p/\mu^2 + 300\log(1/\epsilon)\right] \leq \epsilon.$$

(Here we use the fact that the events $i \in \mathbf{I}$ are $\log(1/\epsilon)$-wise independent.) If that bad event does not occur, then

$$\|w_{\mathbf{I}}\|_2^2 \leq (2p + 300\mu^2\log(1/\epsilon)) \cdot \|w_T\|_2^2,$$

and hence $\|w_{\mathbf{I}}\|_2 \leq O(\sqrt{p} + \mu\sqrt{\log(1/\epsilon)}) \cdot \|w_T\|_2$, completing the proof of Equation (5). In this case, the hypotheses of Proposition 6.5 are satisfied with $\mu' = O(\sqrt{p} + \mu\sqrt{\log(1/\epsilon)})$, so indeed,

$$\begin{aligned}\Pr_{\mathbf{z}}[\Phi\restriction_{\mathbf{I},\mathbf{z}} \text{ is not } \epsilon\text{-close to a constant}] &\leq O\left(\mu\log(1/\epsilon) + \sqrt{p\log(1/\epsilon)}\right) \\ &= \gamma + O\left(\sqrt{p\log(1/\epsilon)}\right). \quad \blacksquare\end{aligned}$$

6.1.3 Restrictions for one layer of an LTF circuit

We would now like to prove that an LTF circuit of super-linear size simplifies under a pseudorandom restriction, with very high probability. More specifically, in this result the $\star$ positions ($\mathbf{I}$) are pseudorandom, but the non-$\star$ positions are filled in using truly random bits ($\mathbf{u}_{[n]\setminus\mathbf{I}}$). This is in contrast to the restrictions we studied in the context of De Morgan formulas, where both components were pseudorandom.

We begin with the special case that each variable in the LTF circuit has a bounded fan-out. (This is analogous to our analysis of De Morgan formulas, where we began with the special case of bounded-read formulas.)

The result below is stated for general parameters. To facilitate its parsing, consider the parametric setting with which we will apply the result: For a very small constant $\delta > 0$, given a circuit with $w = n^{1+\delta}$ wires in which each variable has fan-out at most $t = n^{O(\delta)}$, we apply a restriction that keeps $p = n^{O(\delta)}$ fraction of the variables alive. Our goal is to deduce that the circuit simplifies to an LTF-DT in which each branch queries at most $D = o(p \cdot n)$ variables and $M = n^{O(\delta)}$ LTFs. Indeed, we prove the following more general technical statement:

Proposition 6.7 (pseudorandom restrictions simplify LTF circuits when the variables have bounded fan-out)**.** *Let $n, w, d, t \in \mathbb{N}$ and $p, \epsilon > 0$. Let $\mathbf{I}$ be a k-wise independent p-regular subset of $[n]$ where $k = p^{-5/6}\log(6w/\epsilon)$, sample $\mathbf{u} \in \{0,1\}^n$ uniformly at random, and let $\boldsymbol{\rho} = (\mathbf{I}, \mathbf{u}_{[n]\setminus\mathbf{I}})$. For any depth-$d$ LTF circuit $C\colon \{0,1\}^n \to \{0,1\}$ with at most w wires where every variable has fan-out at most t, there is an $(n, M, D, d-1, w, w)$-LTF-DT $\mathbf{T}$ (jointly distributed with $\boldsymbol{\rho}$) that is consistent with $C\restriction_{\boldsymbol{\rho}}$ such that $\mathbb{E}[\mathsf{Err}(\mathbf{T})] \leq \epsilon$ and*

$$\begin{aligned}D &= O((p^{7/6} \cdot w + p^{-5/2} \cdot t) \cdot \log^2(w/\epsilon)) \\ M &= O(p^{-2/3} \cdot t).\end{aligned}$$

Proof. We first analyze the effect of the restriction $\boldsymbol{\rho}$ on C and prove a number of technical claims, while also explaining intuitively and in advance how we intend to construct an LTF decision tree that approximates $C\restriction_{\boldsymbol{\rho}}$. Afterwards we will rely on these technical claims to formally show how to construct the LTF-DT with error indicators.

Let W be the set of LTF gates Φ such that every input to Φ is a variable (not another gate). We partition $W = W^{\mathtt{H}} \dot\cup W^{\mathtt{L}}$, where $W^{\mathtt{H}}$ is the set of ("heavy") gates with fan-in at least $p^{-5/6}$ and $W^{\mathtt{L}} = W \setminus W^{\mathtt{H}}$. Let $X = \{x_1, \ldots, x_n\}$ be the set of variables.

	notation	our main setting
number of wires	w	$n^{1+\delta}$
fraction of live variables	p	$n^{-O(\delta)}$
target error probability	ϵ	$2^{-O(n^{\delta})}$
heavy gates fan-in	$\geq p^{-5/6}$	$\geq n^{O(\delta)}$
variable fan-out	$\leq t$	$\leq n^{O(\delta)}$

Figure 2: The main parameters in the proof

Light gates. Consider the wires between W^{L} and X. We say that two distinct variables $x_i, x_j \in X$ are a connected pair if there exists $g \in W^{\mathrm{L}}$ such that both x_i and x_j feed into g. The probability that both variables in a connected pair survive is p^2. The number of connected pairs is less than $p^{-5/6} \cdot w$, since each wire participates in fewer than $p^{-5/6}$ connected pairs. Therefore, in expectation, the number of connected pairs that survive the restriction is bounded by $p^2 \cdot p^{-5/6} \cdot w = p^{7/6} \cdot w$. The next claim says that a similar bound holds with high probability.

Claim 6.7.1. *With probability $1 - \epsilon/6$, the number of connected pairs such that both variables in the pair are in $\mathbf{I}$ is $O(p^{7/6} \cdot w + p^{-5/2} \cdot t \cdot \log(w/\epsilon))$.*

Proof. Partition W^{L} into blocks $B_1, \ldots, B_r$, where $r \leq \log(p^{-5/6})$, such that each block B_i consists of all gates with fan-in more than 2^{i-1} and at most 2^i. Fix a block $i \in [r]$. Form a graph with vertex set B_i, where g is adjacent to g' if there is some variable $x_i \in X$ that feeds into both g and g'. Since each gate g has fan-in less than $p^{-5/6}$ and each variable has fan-out at most t, the maximum degree of this graph is less than $p^{-5/6} \cdot t$. Therefore, there is a proper coloring of the graph using $O(p^{-5/6} \cdot t)$ colors.

For each gate $g \in W^{\mathrm{L}}$, let $\mathbf{a}_g \in [0, 1]$ be the fraction of pairs of variables feeding into g such that both variables are in $\mathbf{I}$. Consider one color class S. Within this color class, gates read disjoint sets of variables. Furthermore, each gate reads fewer than $p^{-5/6}$ variables. Therefore, the variables $\mathbf{a}_g$ for $g \in S$ are k'-wise independent where $k' = \log(6w/\epsilon)$. Furthermore, $\mathbb{E}[\mathbf{a}_g] = p^2$, so

$$\mathbb{E}\left[\sum_{g \in S} \mathbf{a}_g\right] = p^2 |S|.$$

By Theorem 4.5, we have

$$\Pr\left[\sum_{g \in S} \mathbf{a}_g \geq 2p^2|S| + 300 \log(6w/\epsilon)\right] \leq \frac{\epsilon}{6w}.$$

Now we shall union bound over all color classes and all blocks. Each color class has at least one gate, and there are at most w gates in total. Therefore, except with probability $\epsilon/6$, for every i, we have

$$\sum_{g \in B_i} \mathbf{a}_g \leq 2p^2|B_i| + O(p^{-5/6} \cdot t \cdot \log(w/\epsilon)).$$

In this case, the number of connected pairs feeding into B_i that survive the restriction is

$$\begin{aligned}\sum_{g\in B_i} \mathbf{a}_g \cdot \binom{\texttt{fan-in}(g)}{2} &\le \binom{2^i}{2} \cdot \sum_{g\in B_i} \mathbf{a}_g \\ &\le \binom{2^i}{2} \cdot (2p^2|B_i| + O(p^{-5/6} \cdot t \cdot \log(w/\epsilon))) \\ &\le O\left(\binom{2^i}{2} \cdot p^{-5/6} \cdot t \cdot \log(w/\epsilon) + p^2 \sum_{g\in B_i} \binom{\texttt{fan-in}(g)}{2}\right),\end{aligned}$$

where the last step holds because $\binom{\texttt{fan-in}(g)}{2} > \binom{2^{i-1}}{2} > \frac{1}{4}\binom{2^i}{2}$ for each $g \in W^{\mathtt{H}}$. Therefore, the total number of connected pairs that survive the restriction is

$$O\left(\left(\sum_{i=1}^{r} \binom{2^i}{2} \cdot p^{-5/6} \cdot t \cdot \log(w/\epsilon)\right) + p^2 \cdot \sum_{g\in W^{\mathtt{L}}} \binom{\texttt{fan-in}(g)}{2}\right).$$

Since the total number of wires is at most w and the fan-in of each $g \in W^{\mathtt{L}}$ is less than $p^{-5/6}$, we have $\sum_{g\in W^{\mathtt{L}}} \binom{\texttt{fan-in}(g)}{2} < p^{-5/6} \cdot w$. Furthermore, $\sum_{i=1}^{r} \binom{2^i}{2} \le 2^{2r+1} \le O(p^{-5/3})$. Therefore the total number of connected pairs that survive is at most

$$O(p^{-5/3} \cdot p^{-5/6} \cdot t \cdot \log(w/\epsilon) + p^2 \cdot p^{-5/6} \cdot w) = O(p^{-5/2} \cdot t \cdot \log(w/\epsilon) + p^{7/6} \cdot w). \square$$

We will condition on the event of Claim 6.7.1, and in this case we will query all of the $O(p^{7/6} \cdot w + p^{-5/2} \cdot t \cdot \log(w/\epsilon))$ variables that participate in surviving connected pairs, such that the only remaining light gates are now projection functions.

Heavy gates. Consider the bipartite subgraph between $W^{\mathtt{H}}$ and X (see Figure 3 for a pictorial representation of this subgraph). We first condition on three events that happen with high probability, and depend only on the choice of $\mathbf{I}$, the set of variables to keep alive. The first event will be that the fan-in of each gate decreases by roughly a factor of p; more accurately:

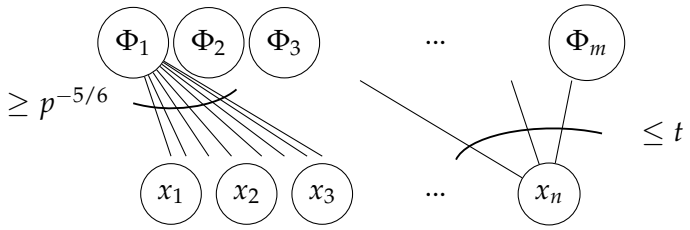

Figure 3: Pictorial representation of the subgraph between $W^{\mathtt{H}}$ and X. Recall that each $\Phi_i \in W^{\mathtt{H}}$ has fan-in at least $p^{-5/6}$, whereas each $x_i \in X$ has fan-out at most t.

Claim 6.7.2. *With probability at least $1 - \epsilon/6$, for each gate $g \in W$, there are at most $2p \cdot \texttt{fan-in}(g) + O(\log(w/\epsilon))$ variables feeding into g that are in $\mathbf{I}$.*

Proof. This follows from Theorem 4.5 and a union bound over the $\le w$ gates in W. $\square$

We now consider two additional high-probability events that depend only on $\mathbf{I}$. Intuitively, we want to claim that after querying an additional small set of variables (on top of the variables assigned by $\boldsymbol{\rho}$), each gate $g \in W^{\mathtt{H}}$ has a decent chance of becoming extremely biased under the uniform random assignment of bits in $\boldsymbol{\rho}$.

In more detail, let $\gamma = p^{1/3}\log(w/\epsilon)$, and let λ be the quantity from Proposition 6.6 with error parameter[20] $\epsilon/(6w)$, so $\lambda = O(p^{-2/3}\cdot\log^2(1/p))$. Let $X^{\texttt{heads}}$ be the set of variables that includes, for each $g \in W^{\texttt{H}}$, the λ variables with the largest weights (in absolute value). Then, we claim that:

Claim 6.7.3. *With probability at least* $1-\epsilon/6$,

$$|\mathbf{I}\cap X^{\texttt{heads}}| \le O(p^{7/6}\cdot w\cdot\log^2(1/p)+\log(1/\epsilon)).$$

Proof. We have $|X^{\texttt{heads}}| \le \lambda\cdot p^{5/6}\cdot w$ (since $|W^{\texttt{H}}| \le p^{5/6}\cdot w$ and each gate in $W^{\texttt{H}}$ contributes at most λ variables to $X^{\texttt{heads}}$). By Theorem 4.5 with $\Delta = p\cdot\lambda\cdot p^{5/6}\cdot w + 300\log(6/\epsilon)$, with probability at least $1-\epsilon/6$, the number of variables in $X^{\texttt{heads}}$ that land in $\mathbf{I}$ is at most

$$O(p\cdot\lambda\cdot p^{5/6}\cdot w+\log(1/\epsilon)) \le O(p^{7/6}\cdot w\cdot\log^2(1/p)+\log(1/\epsilon)).$$

Here we use the fact that $\mathbf{I}$ is $\log(6/\epsilon)$-wise independent. □

Claim 6.7.4. *Let* $\mathbf{I}' = \mathbf{I}\setminus X^{\texttt{heads}}$. *With probability at least* $1-\epsilon/6$ *over the choice of* $\mathbf{I}$, *for every* $g\in W^{\texttt{H}}$,

$$\Pr_{\mathbf{u}\in\{0,1\}^n}[g\restriction_{\mathbf{I}',\mathbf{u}} \text{ is not } (\epsilon/(6w))\text{-close to a constant}] \le O\left(p^{1/3}\log(w/\epsilon)\right). \tag{6}$$

Proof. For each $g\in W^{\texttt{H}}$, by Proposition 6.6, Equation (6) holds with probability $1-\epsilon/(6w)$ over the choice of $\mathbf{I}$ (note that $\sqrt{p\log(1/\epsilon)} < \gamma$.) A union bound over the $\le w$ gates in $W^{\texttt{H}}$ completes the proof. □

Our LTF-DT will query all the variables in $X^{\texttt{heads}}$, and then we will deal with the gates according to the partition in the next claim. We will replace each gate in $\mathbf{W}_1^{\texttt{H}}$ with the corresponding constant; we will query all the variables feeding into the gates in $\mathbf{W}_2^{\texttt{H}}$; we will query the gates in $\mathbf{W}_3^{\texttt{H}}$ themselves.

Claim 6.7.5. *Let* I *be such that the conclusions of Claims 6.7.2 and 6.7.4 are both satisfied. Let* $I' = I\setminus X^{\texttt{heads}}$ *and* $\mathbf{z} = \mathbf{u}_{[n]\setminus I'}$. *With probability* $1-\epsilon/6$ *over* $\mathbf{z}$, $W^{\texttt{H}}$ *can be partitioned into three sets,* $W^{\texttt{H}} = \mathbf{W}_1^{\texttt{H}}\cup\mathbf{W}_2^{\texttt{H}}\cup\mathbf{W}_3^{\texttt{H}}$, *such that under the restriction* $(I',\mathbf{z})$:

1. *Every gate in* $\mathbf{W}_1^{\texttt{H}}$ *is* $(\epsilon/(6w))$*-close to a constant.*
2. *The number of living wires feeding into* $\mathbf{W}_2^{\texttt{H}}$ *is at most* $O(p^{7/6}\cdot w\cdot\log^2(w/\epsilon))$.
3. $|\mathbf{W}_3^{\texttt{H}}| \le O(p^{-2/3}\cdot t)$.

Proof. Initially, partition $W^{\texttt{H}}$ into blocks $B_1,\dots,B_r$, where $r\le\log w$, such that each block B_i consists of all gates with fan-in more than 2^{i-1} and at most 2^i. For each gate $g\in W^{\texttt{H}}$, by Claim 6.7.4, under the restriction $(I',\mathbf{z})$, the probability that g is not $(\epsilon/(6w))$-close to a constant is at most $\varphi = O(p^{1/3}\cdot\log(w/\epsilon))$. We let $\mathbf{W}_3^{\texttt{H}}$ be the union of all gates in blocks B_i with $|B_i| \le \varphi^{-2}\cdot t\cdot\log(6r/\epsilon)$, so indeed,

$$\begin{aligned}|\mathbf{W}_3^{\texttt{H}}| &\le r\cdot\varphi^{-2}\cdot t\cdot\log(6r/\epsilon)\\ &\le O(\log w\cdot p^{-2/3}\cdot\log^{-2}(w/\epsilon)\cdot t\cdot\log(\log(w)/\epsilon))\\ &< O(p^{-2/3}\cdot t).\end{aligned}$$

We let $\mathbf{W}_1^{\texttt{H}}$ be the set of gates in $W^{\texttt{H}}\setminus\mathbf{W}_3^{\texttt{H}}$ that are $(\epsilon/(6w))$-close to a constant after the restriction, and we let $\mathbf{W}_2^{\texttt{H}} = W^{\texttt{H}}\setminus(\mathbf{W}_1^{\texttt{H}}\cup\mathbf{W}_3^{\texttt{H}})$. All that remains is to bound the number of living wires feeding into $\mathbf{W}_2^{\texttt{H}}$.

[20] I.e., the good event in Proposition 6.6 is that the restricted function is $(\epsilon/(6w))$-close to a constant.

Consider some block B_i with $|B_i| > \varphi^{-2} \cdot t \cdot \log(6r/\epsilon)$. For each $g \in B_i$, we have $\Pr[g \in W_2^{\mathtt{H}}] \leq \varphi$. Admittedly, these events are not completely independent, but each variable has fan-out at most t, so we may apply the read-t Chernoff bound (Theorem 4.4) to our uniform choice of $\mathbf{z}$ (recall that I is fixed). This shows that

$$\Pr[|B_i \cap \mathbf{W}_2^{\mathtt{H}}| \geq 2\varphi|B_i|] \leq \exp(-\varphi^2|B_i|/t) < \frac{\epsilon}{6r}.$$

Therefore, by the union bound, with probability $1 - \epsilon/6$ every block B_i contributes at most $2\varphi|B_i|$ gates to $\mathbf{W}_2^{\mathtt{H}}$.

Furthermore, recall that we are assuming that I satisfies the conclusion of Claim 6.7.2, so each gate in B_i has at most $2p \cdot 2^i + O(\log(w/\epsilon))$ living wires. Therefore, in total, the number of living wires feeding into $\mathbf{W}_2^{\mathtt{H}}$ is at most

$$\begin{aligned}\sum_{i=1}^{r} 2\varphi \cdot |B_i| \cdot (2p \cdot 2^i + O(\log(w/\epsilon))) &\leq \varphi \cdot \left(\left(4p \sum_{i=1}^{r} |B_i| \cdot 2^i\right) + O\left(\log(w/\epsilon) \cdot \sum_{i=1}^{r} |B_i|\right)\right) \\ &\leq \varphi \cdot O\left(pw + p^{5/6} \cdot w \cdot \log(w/\epsilon)\right) \\ &\leq O(p^{7/6} \cdot w \cdot \log^2(w/\epsilon)) \,,\end{aligned}$$

where we used the fact that the number of gates in $W^{\mathtt{H}}$ is at most $p^{5/6} \cdot w$. □

Constructing the `LTF-DT` with error indicators. With probability $1 - 4\epsilon/6$ over the choice of restriction $\boldsymbol{\rho}$, the conclusions of Claims 6.7.1, 6.7.2, 6.7.3, and 6.7.4 are all satisfied. If not, we set $\mathbf{T}$ to be a trivial tree with $\mathbf{T}(x) = 0$ and $\mathsf{Err}(\mathbf{T}, x) = 1$ for all x. Assume now that the conclusions of Claims 6.7.1, 6.7.2, 6.7.3, and 6.7.4 are all satisfied.

Let $\mathbf{I}' = \mathbf{I} \setminus X^{\mathtt{heads}}$. On input $x \in \{0,1\}^{\mathbf{I}}$, our tree $\mathbf{T}$ first queries the variables in $X^{\mathtt{heads}}$. These queries, together with $\boldsymbol{\rho}$, define an assignment $\mathbf{z} \in \{0,1\}^{[n] \setminus \mathbf{I}'}$. If this assignment $\mathbf{z}$ does not satisfy the conclusion of Claim 6.7.5, the corresponding node of $\mathbf{T}$ is a leaf labeled by the constant zero function and $\mathcal{E}_\ell = \{1\}$ so that $\mathsf{Err}(\mathbf{T}, x) = 1$.

Assume now that the conclusion of Claim 6.7.5 is satisfied, so we obtain a partition of $W^{\mathtt{H}}$ into $W^{\mathtt{H}} = \mathbf{W}_1^{\mathtt{H}} \cup \mathbf{W}_2^{\mathtt{H}} \cup \mathbf{W}_3^{\mathtt{H}}$. The gates in $\mathbf{W}_1^{\mathtt{H}}$ are close to constants, after restricting according to $(\mathbf{I}', \mathbf{z})$. Let $\widetilde{C\restriction_{\mathbf{I}',\mathbf{z}}}$ be the circuit obtained from $C\restriction_{\mathbf{I}',\mathbf{z}}$ by replacing each gate in $\mathbf{W}_1^{\mathtt{H}}$ with the corresponding constant. Write $\mathbf{W}_1^{\mathtt{H}} = \{\Phi_1, \ldots, \Phi_r\}$ (where $r \leq w$), and let $\Sigma = \{\sigma_1, \ldots, \sigma_r\}$ be the corresponding set of constants. For each $\Phi_i \in \mathbf{W}_1^{\mathtt{H}}$, let

$$\mathbf{E}_i = \begin{Bmatrix} \Phi_i\restriction_{\mathbf{I}',\mathbf{z}} & \sigma_i = 0 \\ 1 - \Phi_i\restriction_{\mathbf{I}',\mathbf{z}} & \sigma_i = 1 \end{Bmatrix},$$

and note that $\mathbf{E}_i(x) = 1$ if and only if $\Phi_i\restriction_{\mathbf{I}',\mathbf{z}}(x) \neq \sigma_i$. Define $\mathcal{E} = \{\mathbf{E}_1, \ldots, \mathbf{E}_r\}$.

Next, $\mathbf{T}$ queries the live variables feeding into $\mathbf{W}_2^{\mathtt{H}}$ and queries each gate in $\mathbf{W}_3^{\mathtt{H}}$. We query also the connected pairs of variables feeding into $W^{\mathtt{L}}$, arriving at a leaf ℓ of $\mathbf{T}$. After replacing each of these variables and gates with the corresponding query value, our circuit $\widetilde{C\restriction_{\mathbf{I}',\mathbf{z}}}$ simplifies to a new `LTF` circuit $\mathbf{C}'$. This new circuit has depth $d - 1$, because for every gate Φ in W, either we queried Φ itself (hence it got replaced with a constant) or we queried all but perhaps one of the input variables to Φ (hence it got replaced with a constant or a literal). We set $C_\ell = \mathbf{C}'$ and $\mathcal{E}_\ell = \mathcal{E}$.

By construction, either $\mathbf{T}(x) = C\restriction_{\boldsymbol{\rho}}(x)$ or else some $\Phi_i(x) \neq \sigma_i$ in which case $\mathsf{Err}(\mathbf{T}, x) \geq 1$, so indeed, $\mathbf{T}$ is consistent with $C\restriction_{\boldsymbol{\rho}}$. Now let us bound $\mathbb{E}[\mathsf{Err}(\mathbf{T})] = \mathbb{E}[\mathsf{Err}(\mathbf{T}, \mathbf{u}_{\mathbf{I}})]$. There is a $4\epsilon/6$ chance that $\boldsymbol{\rho}$ is a bad restriction causing us to immediately halt with $\mathsf{Err}(\mathbf{T}, \mathbf{u}_{\mathbf{I}}) = 1$. The probability that the good event of Claim 6.7.5 does not occur is at most $\epsilon/6$. Each gate Φ_i that we replaced by a constant was $(\epsilon/(6w))$-close to that constant, so

$\mathbb{E}[\mathbf{E}_i(\mathbf{u}_\mathbf{I})] \le \epsilon/(6w)$. There are at most w gates in the circuit, so by linearity of expectation, $\mathbb{E}[\mathsf{Err}(\mathbf{T})] \le 4\epsilon/6 + \epsilon/6 + \epsilon/6 = \epsilon$.

Let us verify that we obtained the claimed parameters for the tree. We query the variables in living connected pairs, the variables in $X^{\mathtt{heads}}$, and the variables feeding into $\mathbf{W}_\mathbf{2}^{\mathtt{H}}$, giving

$$\begin{aligned} D &\le O(p^{7/6} \cdot w + p^{-5/2} \cdot t \cdot \log(w/\epsilon)) \\ &\quad + O(p^{7/6} \cdot w \cdot \log^2(1/p) + \log(1/\epsilon)) \\ &\quad + O(p^{7/6} \cdot w \cdot \log^2(w/\epsilon)) \\ &\le O((p^{7/6} \cdot w + p^{-5/2} \cdot t) \cdot \log^2(w/\epsilon)), \end{aligned}$$

where the last step uses the fact that $\log(1/p) \le \log w$ without loss of generality. (After all, if $p < 1/w$, then our claimed bound on D is greater than w.) Also, we queried the gates from $\mathbf{W}_\mathbf{3}^{\mathtt{H}}$, giving

$$M \le O(p^{-2/3} \cdot t).$$

Finally, note that for each leaf of the tree, none of the variables queried along the path to that leaf appear as input variables in the circuit labeling that leaf. (This is required by Definition 6.1.) ■

We now state a result analogous to Proposition 6.7, but that doesn't assume the variables have bounded fan-out. Intuitively, we will reduce the general case to the case of bounded fan-out by using the decision tree to query "heavy" variables.

Proposition 6.8 (pseudorandom restrictions simplify LTF circuits, even if the variable fan-out is large)**.** *Let $n, w, d \in \mathbb{N}$ and $p, \epsilon > 0$. Let $\mathbf{I}$ be a k-wise independent p-regular subset of $[n]$ where $k = p^{-5/6}\log(6w/\epsilon)$, sample $\mathbf{u} \in \{0,1\}^n$ uniformly at random, and let $\boldsymbol{\rho} = (\mathbf{I}, \mathbf{u}_{[n]\setminus\mathbf{I}})$. For any depth-$d$ LTF circuit $C\colon \{0,1\}^n \to \{0,1\}$ with at most w wires, there is an $(n, M, D, d-1, w, w)$-LTF-DT $\mathbf{T}$ (jointly distributed with $\boldsymbol{\rho}$) that is consistent with $C\restriction_\rho$ such that $\mathbb{E}[\mathsf{Err}(\mathbf{T})] \le \epsilon$ and*

$$\begin{aligned} D &= O((p^{7/6} \cdot w + p^{-11/3}) \cdot \log^2(w/\epsilon)) \\ M &= O(p^{-11/6}). \end{aligned}$$

Proof. Let $X^{\mathtt{H}} \subseteq [n]$ be the set of variables with fan-out more than t, where $t = \lceil p^{-7/6} \rceil$, and let $X^{\mathtt{L}} = [n] \setminus X^{\mathtt{H}}$. For each $z \in \{0,1\}^{X^{\mathtt{H}}}$, define $C_z = C\restriction_{(X^{\mathtt{L}}, z)}$, a depth-$d$ LTF circuit with at most w wires where every variable has fan-out at most t. Applying Proposition 6.7 to C_z gives some tree $\mathbf{T}_z$ (jointly distributed with $\boldsymbol{\rho}$) that is consistent with $(C_z)\restriction_\rho$. On input $x \in \{0,1\}^{\mathbf{I}}$, our tree $\mathbf{T}$ first queries every variable in $X^{\mathtt{H}} \cap \mathbf{I}$. Together with $\boldsymbol{\rho}$, the answers to these queries define some assignment $z \in \{0,1\}^{X^{\mathtt{H}}}$. Our tree $\mathbf{T}$ then simulates $\mathbf{T}_z(x)$. As before (and as required in Definition 6.1), none of the variables queried along the path to a leaf of the tree appear as input variables in the circuit labeling that leaf.

Now let us bound $\mathbb{E}[\mathsf{Err}(\mathbf{T})]$. By construction,

$$\begin{aligned} \mathbb{E}[\mathsf{Err}(\mathbf{T})] &= \mathbb{E}[\mathsf{Err}(\mathbf{T}, \mathbf{u}_\mathbf{I})] \\ &= \mathbb{E}[\mathsf{Err}(\mathbf{T}_{\mathbf{u}_{X^{\mathtt{H}}}}, \mathbf{u}_\mathbf{I})] \\ &= \mathbb{E}\left[\sum_{z \in \{0,1\}^{X^{\mathtt{H}}}} \mathsf{Err}(\mathbf{T}_z, \mathbf{u}_\mathbf{I}) \cdot \mathbb{1}_{z = \mathbf{u}_{X^{\mathtt{H}}}}\right] \\ &= \sum_{z \in \{0,1\}^{X^{\mathtt{H}}}} \mathbb{E}[\mathsf{Err}(\mathbf{T}_z, \mathbf{u}_\mathbf{I}) \cdot \mathbb{1}_{z = \mathbf{u}_{X^{\mathtt{H}}}}]. \end{aligned}$$

For a fixed z, the circuit C_z only reads variables in $X^{\mathtt{L}}$. Therefore, the random variable $\mathsf{Err}(\mathbf{T}_z, \mathbf{u}_{\mathbf{I}})$ is independent of $\mathbf{u}_{X^{\mathtt{H}}}$. Therefore,

$$\mathbb{E}[\mathsf{Err}(\mathbf{T})] = \sum_{z \in \{0,1\}^{X^{\mathtt{H}}}} \mathbb{E}[\mathsf{Err}(\mathbf{T}_z, \mathbf{u}_{\mathbf{I}})] \cdot \Pr[z = \mathbf{u}_{X^{\mathtt{H}}}] = \sum_{z \in \{0,1\}^{X^{\mathtt{H}}}} \mathbb{E}[\mathsf{Err}(\mathbf{T}_z)] \cdot 2^{-|X^{\mathtt{H}}|} \leq \epsilon.$$

Finally, since C has at most w wires, $|X^{\mathtt{H}}| \leq w/t \leq p^{7/6} \cdot w$, hence the number of variable queries made by $\mathbf{T}$ is at most

$$|X^{\mathtt{H}}| + O((p^{7/6} \cdot w + p^{-5/2} \cdot t) \cdot \log^2(w/\epsilon)) = O((p^{7/6} \cdot w + p^{-11/3}) \cdot \log^2(w/\epsilon)),$$

and the number of LTF queries made by $\mathbf{T}$ is at most

$$O(p^{-2/3} \cdot t) = O(p^{-11/6}). \quad \blacksquare$$

6.1.4 Restrictions for one layer of the leaves of an LTF-DT

Recall that given a LTF circuit of depth d and super-linear size, we want to iteratively apply multiple restrictions that will simplify the circuit. Since Proposition 6.7 asserts that the first restriction will already simplify the circuit to an LTF-DT, we now want to show that further restrictions simplify an LTF-DT to a simpler LTF-DT (in particular, one whose leaves are labeled by shallower LTF circuits). This is proved in the following result.

Similarly to Proposition 6.7, the result below is stated for general parameters, and to facilitate its parsing we suggest the following specific setting with which it will be applied. For a very small constant $\delta > 0$, we are given an LTF-DT whose leaves are labeled by circuits with $w = n^{1+\delta}$ wires, and in which each branch queries at most $D = o(n)$ variables and $M = n^{O(\delta)}$ LTF gates. We will apply a restriction that keeps $p = n^{O(\delta)}$ fraction of the variables alive, and want to deduce that the LTF-DT simplifies to one whose leaves are labeled by shallower circuit, such that each branch queries at most $D' = p \cdot D + o(p \cdot n)$ variables and $M' = M + n^{O(\delta)}$ LTFs, with tiny error probability $\epsilon \approx 2^{p^{-.01}}$. These parameter settings are formally stated in Corollary 6.10 below.

In the statement below, instead of using the notation D and D' (and so forth) we will denote the parameters with a subscript d (e.g., D_d and D_{d-1}). We use this notation because we will later use the result in a recursive depth-reduction process, in which all the parameters will change in each depth-reduction step. Nevertheless, we wish to stress that in the generic statement below these parameters are *not* functions of the depth.

Proposition 6.9 (pseudorandom restrictions simplify LTF-DTs)**.** *Let $n, w, d, M_d \in \mathbb{N}$ and $p, \epsilon > 0$. Let $\mathbf{I}$ be a k-wise independent p-regular subset of $[n]$ where $k = p^{-5/6} \cdot (M_d + \log(12w/\epsilon))$, sample $\mathbf{u} \in \{0,1\}^n$ uniformly at random, and let $\boldsymbol{\rho} = (\mathbf{I}, \mathbf{u}_{[n] \setminus \mathbf{I}})$. For any function $C \colon \{0,1\}^n \to \{0,1\}$ and any (n, M_d, D_d, d, w, e_d)-LTF-DT T_d consistent with C, there is an $(n, M_{d-1}, D_{d-1}, d-1, w, e_{d-1})$-LTF-DT $\mathbf{T}_{d-1}$ (jointly distributed with $\boldsymbol{\rho}$) that is consistent with $C{\restriction_{\boldsymbol{\rho}}}$ such that*

$$\begin{aligned}
\mathbb{E}[\mathsf{Err}(\mathbf{T}_{d-1})] &\leq \mathsf{Err}(T_d) + \epsilon \\
D_{d-1} &= O(p \cdot D_d + (p^{7/6} \cdot w + p^{-11/3}) \cdot (M_d^2 + \log^2(w/\epsilon))) \\
M_{d-1} &= M_d + O(p^{-11/6}) \\
e_{d-1} &= e_d + w.
\end{aligned}$$

Before turning to the proof, let us state the special case of Proposition 6.9 that was mentioned above and that we will use for our main result.

Corollary 6.10 (the special case of Proposition 6.9 used in our main result)**.** *Let $\delta > 0$ be sufficiently small and let $n, d \in \mathbb{N}$. For suitable values $p = n^{-O(\delta)}$ and $k = n^{O(\delta)}$, let $\mathbf{I}$ be a k-wise independent p-regular subset of $[n]$. Sample $\mathbf{u} \in \{0,1\}^n$ uniformly at random and let $\boldsymbol{\rho} = (\mathbf{I}, \mathbf{u}_{[n]\setminus\mathbf{I}})$. For any function $C\colon \{0,1\}^n \to \{0,1\}$ and any $(n, M_d, D_d, d, n^{1+\delta}, e_d)$-LTF-DT T_d consistent with C, there is an $(n, M_{d-1}, D_{d-1}, d-1, n^{1+\delta}, e_{d-1})$-LTF-DT $\mathbf{T}_{d-1}$ (jointly distributed with $\boldsymbol{\rho}$) that is consistent with $C\restriction_{\boldsymbol{\rho}}$ such that*

$$\begin{aligned}
\mathbb{E}[\mathsf{Err}(\mathbf{T}_{d-1})] &\le \mathsf{Err}(T_d) + 2^{-\Omega(n^\delta)} \\
D_{d-1} &= O(p \cdot D_d) + n^{-\Omega(\delta)} \cdot p \cdot n \\
M_{d-1} &= M_d + n^{O(\delta)} \\
e_{d-1} &= e_d + n^{1+\delta}.
\end{aligned}$$

Proof of Proposition 6.9. Let L be the leaves of T_d. Let $\ell(x) \in L$ be the leaf reached by $T_d(x)$, and for each $\ell \in L$, let V_ℓ be the set of variables queried by T_d along the path to ℓ. Loosely speaking, we will first show that with high probability, the restriction decreases the length of every path in T_d by a factor of approximately p. Then, we use Proposition 6.7 to simplify the functions that label the leaves of T_d.

Shortening the paths in the tree. On input $x \in \{0,1\}^{\mathbf{I}}$, our new tree $\mathbf{T}_{d-1}$ begins by simulating the tree portion of $T_d(x \circ \boldsymbol{\rho})$, querying x whenever T_d makes a query to a variable in $\mathbf{I}$. However, if $\mathbf{T}_{d-1}$ finds that this process would require making more than q queries to x where $q = 2p \cdot D_d + 300 \log(2/\epsilon)$, then instead of making query number $\lfloor q \rfloor + 1$, our tree $\mathbf{T}_{d-1}$ halts, outputs 0, and sets $\mathcal{E}_\ell = \{1\}$ for that leaf.

Let us bound the contribution from these events to $\mathbb{E}[\mathsf{Err}(\mathbf{T}_{d-1})]$, i.e., the probability that we are forced to halt early in that way with respect to $\boldsymbol{\rho}$ and the uniform random input $\mathbf{u}_{\mathbf{I}}$. Since $\mathbf{u}_{\mathbf{I}} \circ \boldsymbol{\rho} = \mathbf{u}$, we are simulating $T_d(\mathbf{u})$, so the contribution to $\mathbb{E}[\mathsf{Err}(\mathbf{T}_{d-1})]$ is

$$\Pr_{\mathbf{I},\mathbf{u}}[|V_{\ell(\mathbf{u})} \cap \mathbf{I}| > q] = \sum_{\ell \in L} \Pr_{\mathbf{u}}[\ell(\mathbf{u}) = \ell] \cdot \Pr_{\mathbf{I}}[|V_\ell \cap \mathbf{I}| > q].$$

For a fixed ℓ, by Theorem 4.5, we have $\Pr[|V_\ell \cap \mathbf{I}| > q] \le \epsilon/2$. Hence, overall we get the bound

$$\sum_{\ell \in L} \frac{\epsilon}{2} \cdot \Pr_{\mathbf{u}}[\ell(\mathbf{u}) = \ell] = \frac{\epsilon}{2}.$$

Approximating leaves by LTF-DTs. Consider now the case that the simulation of $T_d(x \circ \boldsymbol{\rho})$ successfully reaches a leaf ℓ of T_d (without needing to make more than q queries to x). For each fixed leaf $\ell \in L$, let C_ℓ be the depth-d LTF circuit labeling ℓ, and let $\mathcal{E}_\ell$ be the set of error indicators labeling ℓ. Let $\mathbf{T}^{(\ell)}$ be the tree (jointly distributed with $\boldsymbol{\rho}$) obtained from applying Proposition 6.7 to C_ℓ with error $\epsilon' := \epsilon \cdot 2^{-M_d - 1}$, so $\mathbf{T}^{(\ell)}$ is consistent with $C_\ell\restriction_{\boldsymbol{\rho}}$. (Note that $\mathbf{I}$ has enough independence to apply Proposition 6.7 with this small error value.) Our tree $\mathbf{T}_{d-1}(x)$ simulates $\mathbf{T}^{(\ell)}(x)$. When $\mathbf{T}^{(\ell)}$ finally reaches a leaf labeled by a circuit C' and a set $\mathcal{E}$ of error indicators, then the corresponding leaf of $\mathbf{T}_{d-1}$ is labeled by the same circuit C' and the set $\mathcal{E} \cup \mathcal{E}_\ell$ of error indicators.

In this way, for every x, we have $\mathbf{T}_{d-1}(x) = C\restriction_{\boldsymbol{\rho}}(x)$ or $\mathsf{Err}(\mathbf{T}_{d-1}, x) \ge 1$. Let us bound $\mathbb{E}[\mathsf{Err}(\mathbf{T}_{d-1})] = \mathbb{E}[\mathsf{Err}(\mathbf{T}_{d-1}, \mathbf{u}_{\mathbf{I}})]$. The contribution from the error indicators of T_d is at most $\mathsf{Err}(T_d)$; now we must bound the contribution from the error indicators of $\mathbf{T}^{(\ell)}$. The leaf $\ell \in L$ that we reach is given by $\ell(\mathbf{u}_{\mathbf{I}} \circ \boldsymbol{\rho}) = \ell(\mathbf{u})$.

It is tempting to say that the contribution from $\mathbf{T}^{(\ell(\mathbf{u}))}$ must be at most ϵ', because for each fixed leaf ℓ, $\mathbb{E}[\mathsf{Err}(\mathbf{T}^{(\ell)})] \le \epsilon'$. Unfortunately, we cannot argue the same way we did in

the proof of Proposition 6.8. The reason is that when accounting for the error, the expected contribution of each $\mathbf{T}^{(\ell)}$ (i.e., for a fixed leaf ℓ of T_d) is the number of errors made by $\mathbf{T}^{(\ell)}$ on a random input *conditioned on the input leading the tree to* ℓ in T_d (i.e., conditioned on the event $\ell(\mathbf{u}) = \ell$). If T_d would only query variables, this conditioning would not matter, since $\mathbf{T}^{(\ell)}$ does not depend on the values of variables queried in the path to ℓ. However, T_d also queries M_d LTFs, and hence the conditioning might focus the contribution of $\mathbf{T}^{(\ell)}$ on a subset of inputs on which $\mathbf{T}^{(\ell)}$ is likely to err. Fortunately, as we show below, this conditioning can only increase the contribution of each leaf by 2^{M_d}, and thus overall contribution of $\mathbf{T}^{(\ell(\mathbf{u}))}$ to the error is at most $\epsilon' \cdot 2^{M_d}$.

For each fixed leaf $\ell \in L$, recall that V_ℓ is the set of variables queried in T_d along the path to ℓ, and let $\vec{v}_\ell \in \{0,1\}^{|V_\ell|}$ be the sequence of values of the edges outgoing from the nodes that correspond to V_ℓ along this path. Similarly, let G_ℓ be the set of at most M_d gates queried in T along the path to ℓ, and let $\vec{g}_\ell \in \{0,1\}^{|G_\ell|}$ be the sequence of values of the edges outgoing from the nodes that correspond to G_ℓ along this path. We define $\mathtt{gts}_\ell \colon \{0,1\}^n \to \{0,1\}^{|G_\ell|}$ to be the function $x \mapsto (g(x))_{g \in G_\ell}$, and similarly define $\mathtt{vrs}_\ell(x) = (x_i)_{x_i \in V_\ell}$. Then

$$\begin{aligned}
\mathop{\mathbb{E}}_{\mathbf{I},\mathbf{u}}\left[\mathsf{Err}(\mathbf{T}^{(\ell(\mathbf{u}))}, \mathbf{u}_\mathbf{I})\right] &= \mathop{\mathbb{E}}_{\mathbf{I},\mathbf{u}}\left[\sum_{\ell \in L} \mathsf{Err}(\mathbf{T}^{(\ell)}, \mathbf{u}_\mathbf{I}) \cdot \mathbb{1}_{\ell(\mathbf{u})=\ell}\right] \\
&= \sum_{\ell \in L} \mathop{\mathbb{E}}_{\mathbf{I},\mathbf{u}}\left[\mathsf{Err}(\mathbf{T}^{(\ell)}, \mathbf{u}_\mathbf{I}) \cdot \mathbb{1}_{\mathtt{vrs}_\ell(\mathbf{u})=\vec{v}_\ell} \cdot \mathbb{1}_{\mathtt{gts}_\ell(\mathbf{u})=\vec{g}_\ell}\right] \\
&\le \sum_{\ell \in L} \mathop{\mathbb{E}}_{\mathbf{I},\mathbf{u}}\left[\mathsf{Err}(\mathbf{T}^{(\ell)}, \mathbf{u}_\mathbf{I}) \cdot \mathbb{1}_{\mathtt{vrs}_\ell(\mathbf{u})=\vec{v}_\ell}\right].
\end{aligned}$$

Now, C_ℓ does not read any of the variables in V_ℓ, and hence neither does $\mathbf{T}^{(\ell)}$. Therefore, the random variables $\mathsf{Err}(\mathbf{T}^{(\ell)}, \mathbf{u})$ and $\mathbb{1}_{\mathtt{vrs}_\ell(\mathbf{u})=\vec{v}_\ell}$ are independent, giving us the bound

$$\begin{aligned}
\mathop{\mathbb{E}}_{\mathbf{I},\mathbf{x}}[\mathsf{Err}(\mathbf{T}^{(\ell(\mathbf{u}))}, \mathbf{u}_\mathbf{I})] &\le \sum_{\ell \in L} \mathop{\mathbb{E}}_{\mathbf{I},\mathbf{u}}[\mathsf{Err}(\mathbf{T}^{(\ell)}, \mathbf{u}_\mathbf{I})] \cdot \Pr_{\mathbf{u}}[\mathtt{vrs}_\ell(\mathbf{u}) = \vec{v}_\ell] \\
&= \sum_{\ell \in L} \mathop{\mathbb{E}}_{\boldsymbol{\rho}}[\mathsf{Err}(\mathbf{T}^{(\ell)})] \cdot \Pr_{\mathbf{u}}[\mathtt{vrs}_\ell(\mathbf{u}) = \vec{v}_\ell] \\
&\le \epsilon' \cdot \sum_{\ell \in L} \Pr_{\mathbf{u}}[\mathtt{vrs}_\ell(\mathbf{u}) = \vec{v}_\ell] \\
&= \epsilon' \cdot \mathop{\mathbb{E}}_{\mathbf{u}}[|\{\ell \in L : \mathtt{vrs}_\ell(\mathbf{u}) = \vec{v}_\ell\}|].
\end{aligned}$$

Now consider some fixed string $u \in \{0,1\}^n$. To find all the leaves ℓ satisfying $\mathtt{vrs}_\ell(u) = \vec{v}_\ell$, we can start at the root; when we reach a node that queries a variable, we take the outgoing edge specified by u, and when we reach a node that queries a gate, we choose an outgoing edge arbitrarily. We always make at most M_d such binary choices, so $|\{\ell \in L : \mathtt{vrs}_\ell(u) = \vec{v}_\ell\}| \le 2^{M_d}$. Therefore,

$$\mathop{\mathbb{E}}_{\mathbf{I},\mathbf{u}}[\mathsf{Err}(\mathbf{T}^{(\ell(\mathbf{u}))}, \mathbf{u})] \le 2^{M_d} \cdot \epsilon'.$$

Summing up, overall we get

$$\mathbb{E}[\mathsf{Err}(\mathbf{T}_{d-1})] \le \mathsf{Err}(T_d) + \frac{\epsilon}{2} + 2^{M_d} \cdot \epsilon' = \mathsf{Err}(T_d) + \epsilon.$$

By Proposition 6.8, we have

$$\begin{aligned}
D_{d-1} &\le q + O((p^{7/6} \cdot w + p^{-11/3}) \cdot \log^2(w/\epsilon')) \\
&= O(p \cdot D_d + (p^{7/6} \cdot w + p^{-11/3}) \cdot (M_d + \log(w/\epsilon))^2)
\end{aligned}$$

and $M_{d-1} = M_d + O(p^{-11/6})$. ∎

6.1.5 Iterated restriction: Simplifying the entire circuit

Our goal now is to iteratively apply Proposition 6.9 to a given LTF circuit, in order to simplify it to an LTF-DT whose leaves are labeled by real sums of LTF functions. This is presented in the following result, which is the formal statement of Proposition 2.2, and whose proof mainly involves inductive calculations of the parameters of such iterative applications. As in the previous section, we first state a general result for a broad range of parameters, and afterwards state a corollary that refers to the specific parameters used in our main result.

Proposition 6.11 (iterated restrictions simplify an LTF circuit to a decision tree with LTFs at the leaves)**.** *Let $n, d, w \in \mathbb{N}$ and $p, \epsilon \in (0, 1/2)$. There is a distribution over subsets $\mathbf{I} \subseteq [n]$ with the following properties.*

1. *Let $\boldsymbol{\rho} = (\mathbf{I}, \mathbf{u}_n)$. If $C\colon \{0,1\}^n \to \{0,1\}$ is a depth-d LTF circuit with at most w wires, then there is an $(n, M, D, 1, w, wd)$-LTF-DT $\mathbf{T}$ (jointly distributed with $\boldsymbol{\rho}$) that is consistent with $C\restriction_{\boldsymbol{\rho}}$ such that*

$$D = (p^{1+40^{-(d-1)}} \cdot w + p^{-4}) \cdot 2^{O(d)} \cdot \log^2(w/\epsilon),$$
$$M = O(p^{-2} \cdot d),$$

   and with probability $1 - \epsilon$, we have $\mathsf{Err}(\mathbf{T}) \le \epsilon$.

2. *For each $i \in [n]$, we have $\Pr[i \in \mathbf{I}] \ge p$. Meanwhile, with probability $1 - \epsilon$, we have*

$$|\mathbf{I}| \le 2^d \cdot pn + O(\log(1/\epsilon)).$$

3. *The set $\mathbf{I}$ can be sampled using $\ell = O(p^{-1} \cdot d \cdot \log(w/\epsilon) \cdot \log n)$ truly random bits such that membership in $\mathbf{I}$ can be computed in time $\ell \cdot \mathrm{polylog}(n)$ on a multitape Turing machine.*

Proof. Without loss of generality, assume that $p > 1/n$ (otherwise we can take $\mathbf{I}$ to be a singleton set consisting of a uniform random element of $[n]$). We view the initial circuit C as an $(n, M_d, D_d, d, w, 0)$-LTF-DT $\mathbf{T}_d$ with the trivial parameters $D_d = M_d = 0$. Then, we will repeatedly apply Proposition 6.9 with error parameter ϵ^2/d, giving a sequence of restrictions $\boldsymbol{\rho}_d, \dots, \boldsymbol{\rho}_2 \in \{0, 1, \star\}^n$ and a sequence of trees $\mathbf{T}_{d-1}, \dots, \mathbf{T}_1$, where $\mathbf{T}_i$ is an $(n, M_i, D_i, i, w, w \cdot (d-i))$-LTF-DT, where M_i and D_i will be computed shortly. We emphasize that we have numbered the restrictions so that we start with $\boldsymbol{\rho}_d$ and finish with $\boldsymbol{\rho}_2$. That way, $\boldsymbol{\rho}_i$ is applied to a tree whose leaves are labeled with circuits of depth i. Note also that for convenience, we are thinking of each $\mathbf{T}_i$ as a function of n variables (though possibly fewer than n variables are influential).[21]

Let $A = \sum_{i=2}^{d} 40^{d-i} = (40^{d-1} - 1)/39$ (a normalization factor). For $i = d, \dots, 2$, we choose $\boldsymbol{\rho}_i$ to be a p_i-regular restriction, where

$$p_i \approx p^{40^{d-i}/A}.$$

Specifically, p_i is the smallest power of two such that $p_i \ge p^{40^{d-i}/A}$. For convenience, we extend this definition also to the case $i = d+1$. Define $p_{i\cdots d} = \prod_{j=i}^{d} p_j$. Observe that $p_i \le p_{i+1}$, and indeed $p_i \le 2p_{i+1}^{40}$. Furthermore, $i < d \implies p_i < 2p_{i+1\cdots d}^{39}$.[22] Let $\overline{\boldsymbol{\rho}} = \boldsymbol{\rho}_2 \circ \cdots \circ \boldsymbol{\rho}_d$, so that for each $j \in [n]$,

$$\Pr[j \in (\overline{\boldsymbol{\rho}})^{-1}(\star)] = p_{2\cdots d} \in [p, 2^{d-1} \cdot p].$$

Let $\mathbf{I} = (\overline{\boldsymbol{\rho}})^{-1}(\star)$, so indeed, for each $i \in [n]$, we have $\Pr[i \in \mathbf{I}] \ge p$. Furthermore, since each set $\boldsymbol{\rho}_i^{-1}(\star)$ is k-wise independent for $k > \log(1/\epsilon)$ (this is required by Proposition 6.9

[21] Note that the bounds in the conclusion of Proposition 6.9 do not refer to the number of variables (only to p, ϵ, and the structural parameters of $\mathbf{T}_i$).

[22] Indeed, $p_i < 2 \cdot p^{40^{d-i}/A}$, and $p_{i+1\cdots d} \ge \prod_{j=i+1}^{d} p^{40^{d-j}/A} = p^{(40^{d-i}-1)/(39A)} \ge p^{40^{d-i}/(39A)}$, so $p_i < 2p_{i+1\cdots d}^{39}$.

regardless of M_i and D_i), $(\overline{\boldsymbol{\rho}})^{-1}(\star)$ is also $\log(1/\epsilon)$-wise independent. Therefore, by Theorem 4.5 with $\Delta = 2^{d-1} \cdot pn + 300\log(1/\epsilon)$, with probability $1-\epsilon$, we have

$$|\mathbf{I}| \leq 2^d \cdot pn + 300\log(1/\epsilon).$$

Let $\mathbf{T} = \mathbf{T}_1$. Clearly, $\mathbf{T}$ is an $(n, M_1, D_1, 1, w, wd)$-LTF-DT that is consistent with $C\restriction_\rho$. Furthermore, for any $i \in \{2, \ldots, d\}$ and any fixing of $\boldsymbol{\rho}_d, \ldots, \boldsymbol{\rho}_{i+1}$, Proposition 6.9 guarantees that

$$\mathop{\mathbb{E}}_{\boldsymbol{\rho}_i}[\mathsf{Err}(\mathbf{T}_{i-1})] \leq \mathsf{Err}(\mathbf{T}_i) + \frac{\epsilon^2}{d}.$$

Note that $\mathbf{T}_i$ does not depend on $\boldsymbol{\rho}_i, \ldots, \boldsymbol{\rho}_2$ and similarly $\mathbf{T}_{i-1}$ does not depend on $\boldsymbol{\rho}_{i-1}, \ldots, \boldsymbol{\rho}_2$. Therefore, by averaging, we get

$$\mathop{\mathbb{E}}_{\overline{\boldsymbol{\rho}}}[\mathsf{Err}(\mathbf{T}_{i-1})] \leq \mathop{\mathbb{E}}_{\overline{\boldsymbol{\rho}}}[\mathsf{Err}(\mathbf{T}_i)] + \frac{\epsilon^2}{d}.$$

We start with $\mathsf{Err}(\mathbf{T}_d) = 0$, so inductively we get $\mathbb{E}_{\overline{\boldsymbol{\rho}}}[\mathsf{Err}(\mathbf{T}_1)] < \epsilon^2$, i.e., $\mathbb{E}[\mathbf{T}] < \epsilon^2$. By Markov's inequality, therefore, except with probability ϵ, we have $\mathsf{Err}(\mathbf{T}) < \epsilon$. Now let us compute the parameters M_i and D_i.

Number of LTF queries: By Proposition 6.9, $M_{i-1} \leq M_i + O(p_i^{-11/6})$. Inductively, this shows that

$$M_i = O\left(\sum_{j=i+1}^{d} p_j^{-11/6}\right) < O(d \cdot p_{i+1}^{-2}). \tag{7}$$

In particular, $M_1 < O(dp_2^{-2}) < O(dp^{-2})$ as claimed.

Number of variable queries: By Proposition 6.9,

$$\begin{aligned}
D_{i-1} &= O\left(p_i \cdot D_i + (p_i^{7/6} \cdot w + p_i^{-11/3}) \cdot (M_i^2 + \log^2(w/\epsilon))\right) \\
&< O\left(p_i \cdot D_i + (p_i^{7/6} \cdot w + p_i^{-11/3}) \cdot p_{i+1}^{-4} \cdot (d\log(w/\epsilon))^2\right) && \text{(Equation (7))} \\
&= O\left(p_i \cdot D_i + (p_i^{16/15} \cdot w + p_i^{-113/30}) \cdot (d\log(w/\epsilon))^2\right) && (p_i \leq 2p_{i+1}^{40}) \\
&\leq C \cdot \left(p_i \cdot D_i + (p_d^{1/15} \cdot p_{i\cdots d} \cdot w + p_i^{-4}) \cdot (d\log(w/\epsilon))^2\right) && (6.1)
\end{aligned}$$

for some constant C, where the last inequality uses the fact that $p_i < 2p_{i+1\cdots d}^{39}$ when $i < d$, which implies that $p_i^{1/15} < 2p_{i+1\cdots d} \cdot p_{i+1\cdots d}^{24/15} \leq 2p_{i+1\cdots d} \cdot p_d^{1/15}$. Let us show by backward induction that

$$D_i \leq (2C)^{d-i} \cdot (p_d^{1/15} \cdot p_{i+1\cdots d} \cdot w + p_{i+1}^{-4}) \cdot (d\log(w/\epsilon))^2.$$

This is true in the base case $i = d$. For the inductive step, we have

$$\begin{aligned}
D_{i-1} &< C \cdot \left(p_i \cdot D_i + (p_d^{1/15} \cdot p_{i\cdots d} \cdot w + p_i^{-4}) \cdot (d\log(w/\epsilon))^2\right) && \text{(Eq. (6.1))} \\
&< C \cdot \left(p_i \cdot D_i + (p_d^{1/15} \cdot p_{i\cdots d} \cdot w + p_i^{-4})\right) \cdot (d\log(w/\epsilon))^2 \\
&< C \cdot \left(p_i \cdot (2C)^{d-i} \cdot (p_d^{1/15} \cdot p_{i+1\cdots d} \cdot w + p_{i+1}^{-4}) + (p_d^{1/15} \cdot p_{i\cdots d} \cdot w + p_i^{-4})\right) \cdot (d\log(w/\epsilon))^2 \\
&&& \text{(induction hypothesis)} \\
&< (2C)^{d-(i-1)} \cdot (p_d^{1/15} \cdot p_{i\cdots d} \cdot w + p_i^{-4}) \cdot (d\log(w/\epsilon))^2.
\end{aligned}$$

Now, $p_d^{1/15} \le 2p^{1/(15A)}$, and $\frac{1}{15A} = \frac{39}{15(40^{d-1}-1)} > 40^{-(d-1)}$, so $p_d^{1/15} \le 2p^{40^{-(d-1)}}$. Therefore, indeed,

$$\begin{aligned} D_1 &\le 2^{O(d)} \cdot (p_d^{1/15} \cdot p \cdot w + p_2^{-4}) \cdot \log^2(w/\epsilon) \\ &\le 2^{O(d)} \cdot (p^{1+40^{-(d-1)}} \cdot w + p^{-4}) \cdot \log^2(w/\epsilon). \end{aligned}$$

Sampling cost: We sample each set $\boldsymbol{\rho}_i^{-1}(\star)$ using the algorithm of Claim 4.8. The seed length is

$$\begin{aligned} O\Big(p_i^{-5/6} \cdot (M_i + \log(w/\epsilon)) \cdot \log(n/p_i)\Big) &< O\Big(p_i^{-5/6} \cdot \log(n) \cdot \log(w/\epsilon)\Big) \cdot M_i \\ &< O\Big(p_i^{-5/6} \cdot \log(n) \cdot \log(w/\epsilon)\Big) \cdot (d \cdot p_{i+1}^{-2}) \quad \text{(Eq. (7))} \\ &< O\Big(p_i^{-1} \cdot d \cdot \log(w/\epsilon) \cdot \log n\Big) \ . \qquad (p_i < p_{i+1}^{12}) \end{aligned}$$

Therefore, summing up, the total seed length for sampling $\mathbf{I}$ is

$$\begin{aligned} O\left(\left(\sum_i p_i^{-1}\right) \cdot d \cdot \log(w/\epsilon) \cdot \log n\right) &< O(p_{2^{\dots}d}^{-1} \cdot d \cdot \log(w/\epsilon) \cdot \log n) \\ &= O(p^{-1} \cdot d \cdot \log(w/\epsilon) \cdot \log n) \end{aligned}$$

as claimed. For each i, the time for computing membership in $\boldsymbol{\rho}_i^{-1}(\star)$ is only a $\mathrm{polylog}(n)$ factor larger than the seed length for sampling $\boldsymbol{\rho}_i^{-1}(\star)$, so the same holds for $\mathbf{I}$ altogether. (Note that to compute membership in $\mathbf{I}$, it suffices to compute membership in each $\boldsymbol{\rho}_i^{-1}(\star)$ *in order* starting with $\boldsymbol{\rho}_d$, so this runtime holds even in the multitape Turing machine model.) ■

Let us now state the instantiation of Proposition 6.11 to the parameter values used in our main result:

Corollary 6.12 (iterated restrictions simplify an LTF circuit to a decision tree with LTFs at the leaves)**.** *Let $n, d \in \mathbb{N}$ and let $\delta = \frac{1}{2} \cdot 50^{-d}$. There is a distribution over subsets $\mathbf{I} \subseteq [n]$ such that:*

1. *(**Approximately $p \cdot n \approx n^{0.9}$ live variables.**) For each $i \in [n]$, we have $\Pr[i \in \mathbf{I}] \ge p$, where $p = n^{-(1+\delta)/10}$; and with probability $1 - 2^{-n^\delta}$, we have $|\mathbf{I}| \le 2^{O(d)} \cdot pn$.*
2. *(**Simplifies LTF circuits of super-linear size.**) Let $\boldsymbol{\rho} = (\mathbf{I}, \mathbf{u}_n)$. For any depth-$d$ LTF circuit $C\colon \{0,1\}^n \to \{0,1\}$ with at most $n^{1+\delta}$ wires, with probability $1 - 2^{-n^\delta}$ there is an $(n, M, D, 1, n^{1+\delta}, d \cdot n^{1+\delta})$-LTF-DT $\mathbf{T}$ consistent with $C\restriction_{\boldsymbol{\rho}}$ such that $\mathsf{Err}(\mathbf{T}) \le 2^{-n^\delta}$ and*

$$\begin{aligned} D &= 2^{O(d)} \cdot (p^{1+\Omega(1)} \cdot n) \\ M &= d \cdot n^{1/4}. \end{aligned}$$

Moreover, the set $\mathbf{I}$ can be sampled using $n^{1/10+O(\delta)} \cdot \mathrm{polylog}(n)$ truly random bits and membership in $\mathbf{I}$ can be computed in time $n^{1/10+O(\delta)} \cdot \mathrm{polylog}(n)$ on a multitape Turing machine.

6.2 Fooling LTF decision trees with error indicators

So far, we have shown that an LTF circuit simplifies under pseudorandom restrictions. The simplified model is a decision tree that queries both variables and LTFs, with LTFs and LTF error indicators at the leaves. In this section, we will show how to fool that simplified model,

which mainly amounts to obtaining an improved PRG for *polytopes* (i.e., functions of the form $\texttt{AND} \circ \texttt{LTF}$) in the extremely-low-error regime.

We will in fact construct a PRG for the more general model of $\texttt{ANY} \circ \texttt{LTF}$; that is, we prove Theorem 1.2. The proof outline was described in Section 2.3.3. As mentioned there, our starting point will be the recent result of Kabanets, Koroth, Lu, Myrisiotis, and Oliveira [KKLMO20]. Using their ideas with appropriate communication protocols for LTFs and PRGs for combinatorial rectangles, one can obtain a PRG with seed length $\widetilde{O}(\sqrt{n \cdot s} \cdot \log(1/\epsilon))$. [23] We will get a better seed length of $\widetilde{O}(\sqrt{n \cdot (s + \log(1/\epsilon))})$.

6.2.1 A low-error PRG for arbitrary functions of a bounded number of LTFs

We begin with a general result on fooling functions of the form $p(g_1, \ldots, g_s)$, where p is a polynomial and each g_i has low *communication complexity*. If p is a multivariate real polynomial $p(x) = \sum_a c_a \prod_i x_i^{a_i}$, we define $L_1(p)$ to be the sum of the absolute values of the coefficients of p, i.e., $L_1(p) = \sum_a |c_a|$. Furthermore, we let $\deg(p)$ denote the total degree of p, i.e., $\deg(p) = \max\{\sum_i a_i : c_a \neq 0\}$.

Recall that a *k-dimensional combinatorial rectangle over the alphabet $\{0,1\}^m$* is a function $f : (\{0,1\}^m)^k \to \{0,1\}$ of the form

$$f(x^{(1)}, \ldots, x^{(k)}) = \prod_{i=1}^{k} f_i(x^{(i)}),$$

where $f_i : \{0,1\}^m \to \{0,1\}$.

Theorem 6.13 (low-error PRG for polynomials of low-communication functions)**.** *Let $k, m \in \mathbb{N}$, and let $\mathbf{x}$ be a distribution over $(\{0,1\}^m)^k$ that ϵ-fools k-dimensional combinatorial rectangles over the alphabet $\{0,1\}^m$. Let $f(x) = p(g_1(x), \ldots, g_s(x))$, where $p : \mathbb{R}^s \to [0,1]$ is a polynomial and for each $i \in [s]$, the function $g_i : (\{0,1\}^m)^k \to \{0,1\}$ can be computed by a randomized k-party number-in-hand communication protocol with communication cost R and failure probability $1/3$. Then $\mathbf{x}$ fools f with error*

$$(\epsilon \cdot L_1(p))^{\Omega(1/R)} \cdot 2^{O(\deg(p))}.$$

Theorem 6.13 is similar to the main claim in the analysis in [KKLMO20]. The improvement is that our theorem tolerates communication protocols with constant error, whereas the analysis in [KKLMO20] requires communication protocols with low error due to a union bound. The effect is that we will achieve a smaller value of R, hence a smaller error. Our improvement is based on a beautiful result by Sherstov.

Theorem 6.14 (Making Polynomials Robust to Noise [She13])**.** *Let $p : \{0,1\}^s \to [0,1]$ be a given polynomial. Then for every $\delta > 0$, there is a polynomial $p_{\text{robust}} : \mathbb{R}^s \to \mathbb{R}$ such that*

$$|p(x) - p_{\text{robust}}(x + \eta)| < \delta$$

for all $x \in \{0,1\}^s$ and all $\eta \in [-1/3, 1/3]^s$. Furthermore,

$$\deg(p_{\text{robust}}) \leq O(\deg(p) + \log(1/\delta))$$

and

$$L_1(p_{\text{robust}}) \leq L_1(p) \cdot 2^{O(\deg(p) + \log(1/\delta))}.$$

[23] To see this, let $k = \sqrt{n/s}$. Expand g as a degree-s real polynomial. We compute each monomial $\bar{g}$ by s repetitions of the communication protocol of Viola [Vio15] (following Nisan [Nis93]) for LTFs, which uses $R = O(\sqrt{n \cdot s} \cdot \log(ns/\epsilon))$ communication bits. We then "fool" this protocol with error $\epsilon/2^{O(s)}$ using the PRG for combinatorial rectangles of Gopalan and Yehudayoff [GY20], whose seed length in this parameter setting is dominated by $\widetilde{O}(R) = \widetilde{O}(\sqrt{n \cdot s} \cdot \log(1/\epsilon))$

(The bound on $L_1(p_{\text{robust}})$ is not explicitly stated in Sherstov's work [She13]. We now briefly explain how the L_1 bound can be verified, assuming familiarity with Sherstov's proof [She13]. In the proof of Sherstov's Theorem 5.2, the polynomial he calls p and the quantities he calls d and D satisfy $L_1(p) \leq 2^{O(d+D)}$. It follows that in the proof of his Theorem 6.2, we have $L_1(P) \leq L_1(\phi) \cdot 2^{O(\deg(\phi)+D)}$. He sets $D = O(\deg(\phi) + \log(1/\delta))$, and the polynomial r at the end of his Theorem 6.2 only increases $L_1(p_{\text{robust}})$ by a constant factor. Note that Sherstov's input polynomial ϕ is defined over the domain $\{\pm 1\}^n$ whereas Theorem 6.14 is stated over the domain $\{0,1\}^n$. The justification is that under the input transformation $\{0,1\} \leftrightarrow \{\pm 1\}$ given by $x \leftrightarrow (-1)^x = 1 - 2x$, the degree of a polynomial does not increase, whereas $L_1(p)$ can go up by at most a factor of $2^{O(\deg(p))}$.)

Proof of Theorem 6.13. Let p_{robust} be the robust version of p with $\delta = (\epsilon \cdot L_1(p))^{\Omega(1/R)} \cdot 2^{O(\deg p)}$. On any given input x, any of the functions $g_1(x), \ldots, g_s(x)$ can be computed by a randomized k-party number-in-hand protocol with error $1/3$ and R bits of communication. For $i = 1, \ldots, s$, denote by $q_i(x)$ the acceptance probability of the randomized protocol for $g_i(x)$. We know that if $g_i(x) = 1$, then $q_i(x) \geq 2/3$ and if $g_i(x) = 0$ then $q_i(x) \leq 1/3$. Now, we wish to show that $\mathbf{x}$ fools $p(g_1(x), \ldots, g_s(x))$. By the robustness of p_{robust}, for every x we have

$$|p(g_1(x), \ldots, g_s(x)) - p_{\text{robust}}(q_1(x), \ldots, q_s(x))| \leq \delta, \tag{6.2}$$

and thus it suffices to show that $\mathbf{x}$ fools $p_{\text{robust}}(q_1(x), \ldots, q_s(x))$. For this we analyze each monomial of p_{robust} separately. Write

$$p_{\text{robust}}(z) = \sum_{a : \sum_i a_i \leq \deg(p_{\text{robust}})} c_a \cdot z_1^{a_1} z_2^{a_2} \cdots z_s^{a_s}.$$

Take any formal monomial $z_1^{a_1} z_2^{a_2} \cdots z_s^{a_s}$ with $\sum_i a_i \leq \deg(p_{\text{robust}})$. We wish to show that $\mathbf{x}$ fools the product $q_1(x)^{a_1} \cdots q_s(x)^{a_s}$. Observe that for any x, $q_1(x)^{a_1} \cdots q_s(x)^{a_s}$ can be thought of as the acceptance probability of a protocol that on input x runs the protocol for $g_1(x)$ a_1 times, each with fresh randomness, then the protocol for $g_2(x)$ a_2 times, each with fresh randomness and so on and so forth – eventually taking the AND of the answers of the $\sum_i a_i$ many protocols. This combined protocol is a randomized protocol in the k-party number-in-hand model that uses $R \cdot \deg(p_{\text{robust}})$ communication bits. Denote the value of this protocol on input x and randomness r by $\pi(x, r)$. By design, $\mathbb{E}_{\mathbf{r}}[\pi(x, \mathbf{r})] = q_1(x)^{a_1} \cdots q_s(x)^{a_s}$ for any x.

Now, for every fixed randomness r, the randomized protocol becomes a deterministic protocol with value $\pi_r(x) = \pi(x, r)$ and communication cost at most $R \cdot \deg(p_{\text{robust}})$. We can write $\pi_r(x) = \sum_{z \in A} \pi_{r,z}(x)$, where A is the set of accepting transcripts and $\pi_{r,z}(x)$ indicates whether π has transcript z on input x and randomness r. For fixed r and z, the predicate $\pi_{r,z}$ can be computed by a k-dimensional combinatorial rectangle over the alphabet $\{0,1\}^m$. Since $\mathbf{x}$ fools k-dimensional combinatorial rectangles over the alphabet $\{0,1\}^m$ with error ϵ, it fools π_r with error $\epsilon|A| \leq \epsilon \cdot 2^{R \cdot \deg(p_{\text{robust}})}$. Therefore,

$$\left| \mathop{\mathbb{E}}_{x \sim \mathbf{x}}[\pi_r(x)] - \mathop{\mathbb{E}}_{x \sim \mathbf{u}_n}[\pi_r(x)] \right| \leq \epsilon \cdot 2^{R \cdot \deg(p_{\text{robust}})}.$$

By averaging, we get

$$\left| \mathop{\mathbb{E}}_{x \sim \mathbf{x}, \mathbf{r}}[\pi_{\mathbf{r}}(x)] - \mathop{\mathbb{E}}_{x \sim \mathbf{u}_n, \mathbf{r}}[\pi_{\mathbf{r}}(x)] \right| \leq \epsilon \cdot 2^{R \cdot \deg(p_{\text{robust}})}.$$

And now since $\mathbb{E}_{\mathbf{r}}[\pi_{\mathbf{r}}(x)] = q_1(x)^{a_1} \cdots q_s(x)^{a_s}$ for every x we have

$$\left| \mathop{\mathbb{E}}_{x \sim \mathbf{x}}[q_1(x)^{a_1} \cdots q_s(x)^{a_s}] - \mathop{\mathbb{E}}_{x \sim \mathbf{u}_n}[q_1(x)^{a_1} \cdots q_s(x)^{a_s}] \right| \leq \epsilon \cdot 2^{R \cdot \deg(p_{\text{robust}})}.$$

Summing over all monomials, we get

$$\begin{aligned}
&\left| \mathop{\mathbb{E}}_{x\sim\mathbf{x}}[p_{\text{robust}}(q_1(x),\ldots,q_s(x))] - \mathop{\mathbb{E}}_{x\sim\mathbf{u}_n}[p_{\text{robust}}(q_1(x),\ldots,q_s(x))] \right| \\
&\quad \leq \sum_a |c_a| \cdot \epsilon \cdot 2^{R\cdot \deg(p_{\text{robust}})} \\
&\quad = L_1(p) \cdot \epsilon \cdot 2^{R\cdot O(\deg(p)+\log(1/\delta))} \\
&\quad = \sqrt{L_1(p)\cdot\epsilon}
\end{aligned} \tag{6.3}$$

by a suitable choice of $\delta = (\epsilon \cdot L_1(p))^{\Omega(1/R)} \cdot 2^{O(\deg p)}$.

Overall, from Eq. (6.2) it follows that

$$\left| \mathop{\mathbb{E}}_{x\sim\mathbf{x}}[p(g_1(x),\ldots,p(g_s(x)))] - \mathop{\mathbb{E}}_{x\sim\mathbf{x}}[p_{\text{robust}}(q_1(x),\ldots,q_s(x))] \right| \leq \delta$$

and

$$\left| \mathop{\mathbb{E}}_{x\sim\mathbf{u}_n}[p(g_1(x),\ldots,p(g_s(x)))] - \mathop{\mathbb{E}}_{x\sim\mathbf{u}_n}[p_{\text{robust}}(q_1(x),\ldots,q_s(x))] \right| \leq \delta\,.$$

In Eq. (6.3) we showed that

$$\left| \mathop{\mathbb{E}}_{x\sim\mathbf{u}_n}[p_{\text{robust}}(q_1(x),\ldots,q_s(x))] - \mathop{\mathbb{E}}_{x\sim\mathbf{x}}[p_{\text{robust}}(q_1(x),\ldots,q_s(x))] \right| \leq \sqrt{L_1(p)\cdot\epsilon} < \delta\,,$$

and thus by combining all three inequalities, applying the triangle inequality, we get that $\mathbf{x}$ fools $p(g_1(\cdot), g_2(\cdot), \ldots, g_s(\cdot))$ with error at most 3δ. ∎

In our case, Nisan [Nis93] showed that LTFs have efficient randomized communication protocols, and more recently Viola [Vio15] improved his result:

Theorem 6.15 (Communication complexity of LTFs [Nis93; Vio15])**.** *Let $\Phi\colon \{0,1\}^n \to \{0,1\}$ be an* LTF*. Under any partition of $[n]$ into k parts, there is a δ-error randomized k-party number-in-hand protocol for Φ with communication cost $O(k\cdot\log(k)\cdot\log(n/\delta))$.*

As an immediate corollary of Theorems 6.13 and 6.15, PRGs for combinatorial rectangles fool arbitrary Boolean functions of a bounded number of LTFs. Specifically, denoting by $\texttt{ANY}_s$ the class of all functions $f\colon \{0,1\}^s \to \{0,1\}$, we consider n-bit functions of the form $\texttt{ANY}_s \circ \texttt{LTF}$. We will partition $[n]$ into k parts of m bits, where k and m are *arbitrary* parametric choices (such that $n = k\cdot m$), and claim that a distribution that δ-fools k-dimensional combinatorial rectangles over alphabet $\{0,1\}^m$ also ϵ-fools $\texttt{ANY}_s \circ \texttt{LTF}$; the error parameter ϵ will correspond to δ and to our choices of values for k and m.

Corollary 6.16 (PRGs for combinatorial rectangles fool $\texttt{ANY}\circ\texttt{LTF}$)**.** *Let $n, m, k \in \mathbb{N}$ with $n = mk$. Let $\delta > 0$ and let $\mathbf{x}$ be a distribution over $(\{0,1\}^m)^k$ that δ-fools k-dimensional combinatorial rectangles over the alphabet $\{0,1\}^m$. Let $s \in \mathbb{N}$ and let $f\colon \{0,1\}^n \to \{0,1\}$ be a function in $\texttt{ANY}_s \circ \texttt{LTF}$. Then $\mathbf{x}$ fools f with error $\epsilon = \delta^{\Omega(1/(k\cdot\log(k)\cdot\log(mk)))}\cdot 2^{O(s)}$.*

Proof. For any function $f\colon \{0,1\}^s \to \{0,1\}$, we can write

$$f(x) = \sum_{y\in\{0,1\}^s} f(y)\cdot\left(\prod_{i\in y^{-1}(1)} x_i\right)\left(\prod_{i\in y^{-1}(0)}(1-x_i)\right).$$

This "brute force" expansion shows that $L_1(f) \leq 4^s$. The claim then follows using Theorem 6.13 and Theorem 6.15, while reying on the fact that

$$\epsilon = (\delta\cdot 4^s)^{1/O(k\cdot\log(k)\cdot\log(mk))}\cdot 2^{O(s)} = \delta^{\Omega(1/(k\cdot\log(k)\cdot\log(mk)))}\cdot 2^{O(s)}.$$ ∎

We wish to plug in an explicit PRG for combinatorial rectangles to obtain an explicit PRG for $\texttt{ANY} \circ \texttt{LTF}$. There is a long line of work developing PRGs for combinatorial rectangles [ASWZ96; EGLNV98; LLSZ97; Lu02; GMRTV12; Vio14; GY20; HLV18; Lee19]. The best seed length is by Gopalan and Yehudayoff [GY20].

However, in our setting there is an additional complication: In our proof we will construct a PRG for $\texttt{ANY} \circ \texttt{LTF}$ functions that are applied to some *unknown subset* of $n' < n$ input bits, and we want its seed length to be noticeably smaller than n' rather than only noticeably smaller than n (the difference between the two will be crucial in our particular parameter setting; see the proof of Theorem 6.25 for details).[24] Moreover, we don't only want our PRG to be explicit (i.e., computable in polynomial time), but also want each output bit of the PRG to be computable in time significantly smaller than n. Therefore we consider the following generalization of combinatorial rectangles, which we call "somewhere combinatorial rectangles".

Definition 6.17. *Let $n, m, k \in \mathbb{N}$ with $n \geq mk$. A k-*dimensional somewhere combinatorial rectangle over the alphabet $\{0,1\}^m$ that has n input bits *is a function $f\colon \{0,1\}^n \to \{0,1\}$ such that there exist functions $f_1, \ldots, f_k\colon \{0,1\}^m \to \{0,1\}$ and indices $1 \leq i_1 < i_2 < \cdots < i_{km} \leq n$ such that for each $x \in \{0,1\}^n$, we have*

$$f(x) = f_1(x_{i_1}, \ldots, x_{i_m}) \wedge f_2(x_{i_{m+1}}, \ldots, x_{i_{2m}}) \wedge \cdots \wedge f_k(x_{i_{(k-1)m+1}}, \ldots, x_{i_{km}}).$$

In words, a somewhere combinatorial rectangle applies a combinatorial rectangle to a subset of its n input bits. Fooling somewhere combinatorial rectangles is extra challenging because we do not know in advance which indices $i_1, \ldots, i_{km}$ will be relevant. If we wanted an explicit PRG for somewhere combinatorial rectangles with the smallest possible seed length, we would use Lee's PRG for a more general model called "product tests" [Lee19]. For ease of analysis, we will instead use the Forbes-Kelley PRG [FK18] (which fools an even more general model) because we can afford the slightly inferior seed length. We now verify that each output bit of the Forbes-Kelley PRG can be computed quickly on a multitape Turing machine.

Theorem 6.18 (Strongly explicit PRGs for combinatorial rectangles [FK18])**.** *For every $n, m \in \mathbb{N}$ and every $\epsilon > 0$, there is a PRG that ϵ-fools somewhere combinatorial rectangles (of any dimension) over the alphabet $\{0,1\}^m$ that have n input bits. The PRG has seed length*

$$O((m + \log(n/\epsilon)) \cdot \log^2 n)$$

and each output bit is computable in time $(m + \log(1/\epsilon)) \cdot \operatorname{polylog}(n)$ on a multitape Turing machine.

Theorem 6.18 follows from Forbes and Kelley's work [FK18] without any new insights, but to verify the theorem we will have to review some of their analysis [FK18]. Recall that a *width-w length-n read-once branching program* (ROBP) is a directed graph consisting of $n+1$ layers $V_0, \ldots, V_n$, each with w vertices. Each vertex has two outgoing edges leading to the next layer labeled 0 and 1 (except the vertices in the last layer, which have no outgoing edges). The program starts at a designated start vertex in V_0 and reads an input $x \in \{0,1\}^n$ to walk through the graph, using x_i to choose the outgoing edge in $V_{i-1} \times V_i$. Finally, the program accepts or rejects x depending on which vertex it arrives at in V_n, thereby defining a function $f\colon \{0,1\}^n \to \{0,1\}$.

[24] The reason that we refer to the subset of n' as "unknown" is that we want each of the n output bits of our PRG to be computable in time $o(n)$, in order to use this PRG as part of our final PRG construction in which each output bit is computable in time $o(n)$. In particular, when computing each output bit we will not have enough time to go over certain previous choices of the PRG that determine which subset of n' variables is the relevant one on which the $\texttt{ANY} \circ \texttt{LTF}$ function is defined.

Forbes and Kelley present an explicit PRG for ROBPs.[25]. Actually, they present *two* closely related PRG constructions (see Section 4 of their paper). The first is a relatively simple "warm-up" construction, and the second is a more refined construction that has a better seed length in some cases. Both constructions can be viewed as being based on pseudorandom restrictions. There are two differences between the two constructions: (a) the restrictions in the warm-up construction are based on t-wise independence whereas the restrictions in the refined construction are based on small-bias distributions and almost t-wise independence, and (b) after performing several restrictions, the warm-up construction fills in any remaining stars with 1, whereas the more refined construction fills in any remaining stars using a t-wise independent distribution.

We will use a "hybrid" of these two constructions: We only use t-wise independence (rather than also small-bias distributions) but we fill-in remaining stars with t-wise independence instead of 1's. The reason we do so is in order to preserve algorithmic simplicity (only using t-wise independence) while still getting a sufficiently good seed length.[26] This yields the following PRG:

Lemma 6.19 (A version of the Forbes-Kelley PRG [FK18])**.** *Let $w, n \in \mathbb{N}$ and $\epsilon > 0$, and let $r = \lceil \log n \rceil$. For a suitable value $t = O(\log(wn/\epsilon))$, let $\boldsymbol{\rho}_1, \ldots, \boldsymbol{\rho}_r \in \{0,1,\star\}^n$ be independent restrictions, each of which is $(1/2)$-regular and t-wise independent. Let $\overline{\boldsymbol{\rho}} = \boldsymbol{\rho}_r \circ \cdots \circ \boldsymbol{\rho}_1$, let $\mathbf{y} \in \{0,1\}^n$ be a t-wise independent string, and let $\mathbf{z} = \mathbf{y} \circ \overline{\boldsymbol{\rho}}$. Then $\mathbf{z}$ fools width-w length-n ROBPs with error ϵ.*

Proof sketch. By [FK18, Lemma 6.3], if $\boldsymbol{\rho} \in \{0,1,\star\}^n$ is a $(1/2)$-regular t-wise independent restriction, then the distribution $\mathbf{u}_n \circ \boldsymbol{\rho}$ fools width-w length-n ROBPs with error $nw \cdot 2^{-\Omega(t)}$ (this is the main step of the proof).[27] In words, informally, applying $\boldsymbol{\rho}$ *preserves the expectation* of the branching program to within a small additive error. The class of Boolean functions on n bits that are computable by width-w ROBPs is closed under restriction, so it follows by induction on r that the distribution $\mathbf{u}_n \circ \overline{\boldsymbol{\rho}}$ fools width-w length-n ROBPs with error $rnw \cdot 2^{-\Omega(t)}$. (Note that $\overline{\boldsymbol{\rho}}$ has a much smaller $\star$-probability than $\boldsymbol{\rho}$, and therefore the number of truly random bits needed to complete $\overline{\boldsymbol{\rho}}$ into an n-bit string is smaller.)

The composition $\overline{\boldsymbol{\rho}}$ is a (2^{-r})-regular t-wise independent restriction. This implies that $\mathbb{E}[|(\overline{\boldsymbol{\rho}})^{-1}(\star)|] = 2^{-r}n \leq 1$. We now think of the coordinates as being merely $((t-1)/300)$-wise independent, which allows us to apply Theorem 4.5 with $\Delta = t - 1$ and conclude that except with probability $2^{-(t-1)/300}$, we have $|(\overline{\boldsymbol{\rho}})^{-1}(\star)| \leq t$. Since $\mathbf{y}$ is t-wise independent, whenever the event $|(\overline{\boldsymbol{\rho}})^{-1}(\star)| \leq t$ happens, for any function $f\colon \{0,1\}^n \to \{0,1\}$ it holds that $\mathbf{y}$ *perfectly* fools $f\restriction_{\overline{\boldsymbol{\rho}}}$. It follows that for any function $f\colon \{0,1\}^n \to \{0,1\}$ we have

$$|\,\mathbb{E}[f(\mathbf{u}_n \circ \overline{\boldsymbol{\rho}})] - \mathbb{E}[f(\mathbf{y} \circ \overline{\boldsymbol{\rho}})]| \leq 2^{-(t-1)/300}.$$

Therefore, $\mathbf{z}$ fools width-w length-n ROBPs with error $rnw \cdot 2^{-\Omega(t)} + 2^{-(t-1)/300}$, which is at most ϵ, provided we use a suitable $t = O(\log(wn/\epsilon))$. ∎

[25] Compared to classic PRGs for space-bounded computation [Nis92; INW94], the main feature of the Forbes-Kelley PRG is that it fools ROBPs that read their input bits *in any order* For us, this feature is not actually necessary, because somewhere combinatorial rectangles can be computed by ROBPs that read their bits in the standard order. We are instead taking advantage of the *simplicity and computational efficiency* of the Forbes-Kelley PRG.

[26] Specifically, we fill in the remaining stars pseudorandomly because this way, we merely need to apply $O(\log n)$ restrictions rather than $O(\log(n/\epsilon))$, giving us a total seed length that depends linearly on $\log(1/\epsilon)$.

[27] The original statement in [FK18, Lemma 6.3] does not explicitly mention the distribution $\mathbf{u}_n \circ \boldsymbol{\rho}$, but refers to the distribution $\mathbf{d} \oplus (\mathbf{t} \wedge \mathbf{u}_n)$, where both $\mathbf{d}$ and $\mathbf{t}$ are t-wise independent n-bit random variables, and $\oplus$ and $\wedge$ denote the bit-wise $\oplus$ and $\wedge$ operations, respectively. To see that the two distributions are identical, think of the original distribution from [FK18] as follows: The choice of $\mathbf{t}$ determines a k-wise independent set of variables $\mathbf{I} \subseteq [n]$, and the choice of $\mathbf{d}$ assigns values for variables outside $\mathbf{I}$ in a t-wise independent manner; then, conditioned on any choices for $\mathbf{d}$ and $\mathbf{t}$, the choice of $\mathbf{u}_n$ assigns truly random values for the variables in $\mathbf{I}$.

Proof of Theorem 6.18. For any k, a k-dimensional somewhere combinatorial rectangle over the alphabet $\{0,1\}^m$ that has n input bits can be computed by a read-once branching program (ROBP) of width $w = 2^m + 1$ and length n. We will now verify that each output bit of the PRG described in Lemma 6.19 can be computed in time $(\log(w/\epsilon)) \cdot \mathrm{polylog}(n)$ on a multitape Turing machine. A similar analysis was performed by Cheraghchi, Hirahara, Myrisiotis, and Yoshida [CHMY21], but they looked at circuit size whereas we are looking at uniform time complexity.

We obtain the string $\mathbf{y}$ and each restriction $\boldsymbol{\rho}_i$ using the algorithm of Claim 4.8. The total seed length is therefore $O(r \cdot k \cdot \log n) = O(\log(wn/\epsilon) \cdot \log^2 n)$. Furthermore, each coordinate of each $\boldsymbol{\rho}_i$ can be computed in time $k \cdot \mathrm{polylog}(n)$, so each coordinate of $\mathbf{z}$ can be computed in time $r \cdot k \cdot \mathrm{polylog}(n) = \log(w/\epsilon) \cdot \mathrm{polylog}(n)$. (Just like in the proof of Proposition 6.11, to compute a coordinate of $\mathbf{z}$, it suffices to compute the corresponding coordinate of $\mathbf{y}$ and of each $\boldsymbol{\rho}_i$ *in order* starting with $\boldsymbol{\rho}_1$, so this runtime bound holds even in the multitape Turing machine model.) ∎

By plugging the PRG of Theorem 6.18 into Corollary 6.16, we will obtain our improved low-error PRG for $\mathtt{ANY} \circ \mathtt{LTF}$. We get a better seed length for functions that ignore most of their input bits.

Corollary 6.20 (low-error PRG for $\mathtt{ANY} \circ \mathtt{LTF}$)**.** *For any $n, n', s \in \mathbb{N}$ and any $\epsilon > 0$, there is an ϵ-PRG for $\mathtt{ANY}_s \circ \mathtt{LTF}$ functions $f\colon \{0,1\}^n \to \{0,1\}$ that ignore all but n' of the input bits. The seed length and the time to compute each output bit on a multitape Turing machine are both*

$$\sqrt{n' \cdot (s + \log(1/\epsilon))} \cdot \mathrm{polylog}(n). \tag{8}$$

Proof. We will use a δ-PRG for k-dimensional somewhere combinatorial rectangles over the alphabet $\{0,1\}^m$ that have n input bits, where $k = n'/m$ and the parameters δ and m will be chosen later. Consider any function $f\colon \{0,1\}^n \to \{0,1\}$ of the form $\mathtt{ANY}_s \circ \mathtt{LTF}$ that ignores all but n' of its n input bits. By applying Corollary 6.16 to the n' relevant bits, we see that for $\delta = (\epsilon \cdot 2^{-O(s)})^{O(k \cdot \log(k) \cdot \log(n'))}$, we fool f with error ϵ. Plugging into Theorem 6.18, the seed length is

$$O((m + \log(n/\delta)) \log^2 n) = (m + k \cdot (s + \log(1/\epsilon))) \cdot \mathrm{polylog}(n).$$

To balance the two terms, we choose $m = \sqrt{n' \cdot (s + \log(1/\epsilon))}$, so $k = \sqrt{n'/(s + \log(1/\epsilon))}$, giving the desired seed length. ∎

6.2.2 Fooling LTF decision trees with error indicators

We now show our low-error PRG for LTF-DT's when the "circuits" at the leaves have depth 1 (i.e., are simply LTFs). In fact, our PRG also fools such LTF-DT's that have *error indicators*, which means that the PRG fools the integer-valued function $T(x) \pm \mathsf{Err}(T, x)$. The proof amounts to observing that our PRG for $\mathtt{ANY}_s \circ \mathtt{LTF}$ with sufficiently low error fools such trees. Once again, we get a better seed length if the tree ignores most of the input bits.

Theorem 6.21 (low-error PRG for LTF-DT's)**.** *For all $n, n', M, D, w, e, \epsilon$, there is a ϵ-PRG for all functions $f\colon \{0,1\}^n \to \mathbb{Z}$ of the form $f(x) = T(x) + \xi \cdot \mathsf{Err}(T, x)$, where $\xi \in \{\pm 1\}$ and T is an $(n, M, D, 1, w, e)$-LTF-DT that ignores all but n' of its input bits. The seed length and the time to compute each output bit on a multitape Turing machine are both*

$$\sqrt{n' \cdot (D + M + \log(e+1) + \log(1/\epsilon))} \cdot \mathrm{polylog}(n).$$

Proof. Let L be the set of leaves. For each leaf $\ell \in L$, let $g_\ell(x) = 1$ if and only if the tree reaches ℓ when it reads x. Let $\Phi_{\ell,0}$ be the LTF labeling ℓ, and let $\mathcal{E} = \{\Phi_{\ell,1}, \dots, \Phi_{\ell,e}\}$ be the set of error indicators labeling ℓ. That way,

$$f(x) = \sum_{\ell \in L} g_\ell(x) \cdot \left(\Phi_{\ell,0}(x) + \xi \cdot \sum_{i=1}^{e} \Phi_{\ell,i}(x) \right).$$

The function g_ℓ is a conjunction of at most M LTFs and at most D literals. Any conjunction of literals can be computed by a single LTF, so the function $g_\ell(x) \cdot \Phi_{i,\ell}(x)$ can be computed by a conjunction of at most $M+2$ LTFs. Therefore, it is δ-fooled by the δ-PRG from Corollary 6.20 with $s = M+2$. Therefore, the same PRG fools f with error $\delta \cdot |L| \cdot (e+1) \le \delta \cdot 2^{D+M} \cdot (e+1)$. Therefore, we should choose $\delta = \epsilon \cdot 2^{-D-M}/(e+1)$, giving the claimed seed length. ∎

6.2.3 Improved low-error PRGs for De Morgan formulas with LTFs at the leaves

As another application of Theorem 6.13, we obtain an improved low-error PRG for size-s *De Morgan formulas* with LTFs at the leaves. Kabanets, Koroth, Lu, Myrisiotis, and Oliveira [KKLMO20] achieved seed length

$$\widetilde{O}(n^{1/2} s^{1/4} \log(1/\epsilon)) \tag{9}$$

for this class, which is trivial if $\log(1/\epsilon) > \sqrt{n}$. Our seed length is

$$\widetilde{O}(n^{1/2} s^{1/4} \log^{1/4}(1/\epsilon) + n^{1/2} \log^{1/2}(1/\epsilon)), \tag{10}$$

which remains nontrivial provided $s \log(1/\epsilon) \ll n^2$ and $\log(1/\epsilon) \ll n$. The second requirement is unavoidable, since fooling conjunctions of literals already requires seed length $\Omega(\log(1/\epsilon))$. When, e.g., $s = n$ and $\epsilon = 2^{-\sqrt{n}}$, both the seed length by [KKLMO20] (Eq. 9) and our seed length for fooling arbitrary functions of LTFs (Eq. 8) are larger than n, whereas our seed length for De Morgan formulas of LTFs is $\widetilde{O}(n^{7/8})$.

To achieve our seed length, we use both Theorem 6.13 and an improved, optimal bound on the low-error approximate degree of a size-s De Morgan formula. Since this result is not necessary for our main constructions, we defer the proof to Appendix A.

6.2.4 Fooling LTF circuits of unbounded depth

As another corollary of Corollary 6.20, we show a PRG for LTF circuits of unbounded depth with at most $O(n/\mathrm{polylog}(n))$ gates.

Corollary 6.22 (PRG for LTF circuits with few gates)**.** *For any $n, s \in \mathbb{N}$ and any $\epsilon > 0$, there is a $\mathrm{poly}(n)$-time computable ϵ-PRG for LTF circuits with s gates with output length n and seed length*

$$\widetilde{O}\left(\sqrt{n \cdot (s + \log(1/\epsilon))} \right).$$

Proof. Let C be an LTF circuit with s gates. The input variables to C are $x_1, \dots, x_n$. Let $y_1, \dots, y_s$ denote the values output by the gates of C in topological order, so y_s is the output value of the circuit, and for each $i \in [s]$, there is some LTF Φ_i such that $y_i = \Phi_i(x_1, \dots, x_n, y_1, \dots, y_{i-1})$.

The proof uses the "guess and verify" technique. For each string $z \in \{0,1\}^s$ (the "guess"), define $g_z(x) = 1$ if and only if during the computation of $C(x)$, for every i, we have $y_i = z_i$ (i.e., verifying the guess). That way,

$$C(x) = \sum_{\substack{z \in \{0,1\}^s \\ z_s = 1}} g_z(x). \tag{11}$$

The key claim is that for a fixed z,

$$g_z(x) = 1 \iff \forall i,\ \Phi_i(x_1,\ldots,x_n,z_1,\ldots,z_{i-1}) = z_i. \tag{12}$$

Indeed, if $g_z(x) = 1$, then $\Phi_i(x_1,\ldots,x_n,z_1,\ldots,z_{i-1}) = \Phi_i(x_1,\ldots,x_n,y_1,\ldots,y_{i-1}) = y_i = z_i$. Conversely, if the right-hand side of Equation (12) holds, then induction on i shows that $z_i = y_i$ for every $i \in [s]$.

Equation (12) implies that $g_z \in \mathtt{AND}_s \circ \mathtt{LTF}$. Therefore, by Equation (11), any δ-PRG for $\mathtt{AND}_s \circ \mathtt{LTF}$ fools C with error $\delta \cdot 2^{s-1}$. Therefore, we may use the PRG of Corollary 6.20 with error $\delta = \epsilon \cdot 2^{-s+1}$. ∎

6.3 The final PRG construction

In this section we combine the restriction procedure from Section 6.1 and the PRG from Section 6.2 to get a PRG for $\mathtt{LTF}$ circuits of super-linear size. As explained in Section 2, we will use the PRG framework of Ajtai and Wigderson [AW85]. Let us now formally state the abstract version of their framework (see [ST19] for a discussion and proof of a nearly identical formulation).

Definition 6.23 (C-to-$\mathsf{C}_{\mathtt{simple}}$ simplifying restriction)**.** *Let $n \in \mathbb{N}$ and $p > 0$. Let C and $\mathsf{C}_{\mathtt{simple}}$ be classes of functions $C\colon \{0,1\}^n \to \{0,1\}$, let $\mathbf{I}$ be a distribution over subsets of $[n]$, and let $\eta > 0$. We say that $\mathbf{I}$ is a* C-to-$\mathsf{C}_{\mathtt{simple}}$ simplifying p-restriction with error η *if:*

1. *For each $i \in [n]$, we have $\Pr[i \in \mathbf{I}] \geq p$.*
2. *For every $C \in \mathsf{C}$, we have*
$$\Pr_{\mathbf{I},\mathbf{z}\in\{0,1\}^{[n]\setminus\mathbf{I}}}\left[C\restriction_{\mathbf{I},\mathbf{z}} \notin \mathsf{C}_{\mathtt{simple}}\right] \leq \eta.$$

Theorem 6.24 (the PRG framework of [AW85]; see [ST19, Theorem 5.1])**.** *Let $n \in \mathbb{N}$, let $p, \epsilon > 0$, and let $\eta = \frac{\epsilon \cdot p}{4\ln(2n/\epsilon)}$. Let C and $\mathsf{C}_{\mathtt{simple}}$ be classes of functions $C\colon \{0,1\}^n \to \{0,1\}$ where C is closed under restriction. Assume that:*

1. *A C-to-$\mathsf{C}_{\mathtt{simple}}$ simplifying p-restriction $\mathbf{I}$ with error η can be sampled using s truly random bits such that membership in $\mathbf{I}$ can be computed in time t on a multitape Turing machine.*
2. *There is an η-PRG for $\mathsf{C}_{\mathtt{simple}}$ with seed length s, where each output bit can be computed in time t on a multitape Turing machine.*

Then there is an ϵ-PRG for C with seed length $O(s \cdot p^{-1} \cdot \log(n/\epsilon))$, where each output bit can be computed in time $O(t \cdot p^{-1} \cdot \log(n/\epsilon))$ on a multitape Turing machine.[28]

We now state our main result.

Theorem 6.25 (PRG for super-linear sized $\mathtt{LTF}$ circuits)**.** *Let $d\colon \mathbb{N} \to \mathbb{N}$ be a function, let $\delta(n) = \frac{1}{4} \cdot 50^{-d(n)}$, and assume that $d(n)$ can be computed in time $O(n^{1-\delta(n)})$ on a multitape Turing machine. There exists an ϵ-PRG for $\mathtt{LTF}$ circuits of depth $d(n)$ with at most $n^{1+\delta(n)}$ wires, whose seed length is $O(n^{1-\delta(n)})$ and whose error is $\epsilon = 2^{-n^{\delta(n)}}$, such that each output bit of the PRG can be computed in time $O(n^{1-\delta(n)})$ on a multitape Turing machine.*

[28] As explained in Section 4, our usual convention is that algorithms receive "all relevant parameter values" as inputs. In this case, the convention might seem to suggest that the algorithms somehow receive descriptions of C and $\mathsf{C}_{\mathtt{simple}}$ as inputs, but that is not what we have in mind. Instead the restriction sampling algorithm merely gets n, p, η, and its seed as inputs; the PRG for $\mathsf{C}_{\mathtt{simple}}$ gets n, η and its seed; the PRG for C gets n, p, ϵ, and its seed. Implicitly, we are assuming that there is an infinite family of pairs $(\mathsf{C}, \mathsf{C}_{\mathtt{simple}})$ – one for each triple (n, p, ϵ).

Proof. For convenience, when n is clear from context we will denote $d = d(n)$, $\delta = \delta(n)$, $\epsilon = \epsilon(n)$, etc. Let $w = n^{1+\delta}$, $p = w^{-1/10}$, and $\eta = \frac{\epsilon \cdot p}{4\ln(2n/\epsilon)}$. Let C be the set of functions $C\colon \{0,1\}^n \to \{0,1\}$ that can be computed by depth-d $\mathtt{LTF}$ circuits with at most w wires, and note that C is closed under restriction. Let $\mathsf{C}_{\mathtt{simple}}$ be the set of functions $C\colon \{0,1\}^n \to \{0,1\}$ such that (a) C ignores all but n' of its n input bits and (b) there is some $(n, M, D, 1, w, wd)$-$\mathtt{LTF\text{-}DT}$ consistent with C, where $\mathsf{Err}(T) \le \eta/2$ and n', M, and D are the parameters when applying Proposition 6.11 with error η, namely

$$\begin{aligned} n' &= 2^d \cdot pn + O(\log(1/\eta)) \le O(2^d \cdot pn) \\ D &= (p^{1+40^{-(d-1)}} \cdot w + p^{-4}) \cdot 2^{O(d)} \cdot \log^2(w/\eta) = p^{1+40^{-(d-1)}} \cdot w \cdot 2^{O(d)} \cdot \log^2(w/\eta) \\ M &= O(p^{-2} \cdot d) \ll D. \end{aligned}$$

Observe that every such C is $(\eta/2)$-upper-sandwiched by $T(x) + \mathsf{Err}(T, x)$ and $(\eta/2)$-lower-sandwiched by $T(x) - \mathsf{Err}(T, x)$.

With these definitions, Proposition 6.11 gives a C-to-$\mathsf{C}_{\mathtt{simple}}$ simplifying p-restriction with error η. The seed length is

$$O(p^{-1} \cdot d \cdot \log(w/\eta) \cdot \log n),$$

and the time to compute membership is $p^{-1} \cdot d \cdot \log(w/\eta) \cdot \mathrm{polylog}(n)$. Recall that Theorem 6.21 gives an $(\eta/2)$-PRG for all functions f of the form $f(x) = T(x) + \zeta \cdot \mathsf{Err}(T, x)$, where $\zeta \in \{\pm 1\}$ and T is an $(n, M, D, 1, w, wd)$-$\mathtt{LTF\text{-}DT}$ that ignores all but n' of its input bits, with the seed length and the time to compute each output bit both bounded by

$$\begin{aligned} &\sqrt{n' \cdot (D + M + \log(wd) + \log(1/\eta))} \cdot \mathrm{polylog}(n) \\ &\quad = \sqrt{n' \cdot p^{1+40^{-(d-1)}} \cdot w} \cdot 2^{O(d)} \cdot \log(w/\eta) \cdot \mathrm{polylog}(n) \\ &\quad \le \widetilde{O}\left(p^{1+\frac{1}{2}40^{-(d-1)}} \cdot w \cdot 2^{O(d)} \cdot \log(1/\epsilon)\right). \end{aligned}$$

The following standard fact implies that the same PRG η-fools $\mathsf{C}_{\mathtt{simple}}$.

Fact 6.25.1. *If a function $f\colon \{0,1\}^n \to \{0,1\}$ is $(\eta/2)$-upper-sandwiched and $(\eta/2)$-lower-sandwiched by functions from a class $\mathcal{F} \subseteq \{\{0,1\}^n \to \mathbb{R}\}$, then any distribution $\mathbf{w}$ over $\{0,1\}^n$ that is $(\eta/2)$-pseudorandom for $\mathcal{F}$ is η-pseudorandom for f.*

Proof. Denoting the lower-sandwiching function by $f^{(low)}$, we have that $\mathbb{E}[f(\mathbf{u}_n)] \le \mathbb{E}[f^{(low)}(\mathbf{u}_n)] + \eta/2 \le \mathbb{E}[f^{(low)}(\mathbf{w})] + \eta \le \mathbb{E}[f(\mathbf{w})] + \eta$, and similarly, using the upper-sandwiching function $f^{(up)}$, we have that $\mathbb{E}[f(\mathbf{u}_n)] \ge \mathbb{E}[f^{(up)}(\mathbf{u}_n)] - \eta/2 \ge \mathbb{E}[f^{(up)}(\mathbf{w})] - \eta \ge \mathbb{E}[f(\mathbf{w})] - \eta$. □

Applying Theorem 6.24 gives us our ϵ-PRG for C. The seed length and the time to compute each output bit (dominated by the PRG for $\mathsf{C}_{\mathtt{simple}}$) are both bounded by

$$\begin{aligned} &\widetilde{O}\left(p^{1+\frac{1}{2}40^{-(d-1)}} \cdot w \cdot 2^{O(d)} \cdot \log(1/\epsilon) \cdot p^{-1} \cdot \log(n/\epsilon)\right) \\ &\quad = \widetilde{O}\left(p^{\frac{1}{2}40^{-(d-1)}} \cdot w \cdot 2^{O(d)} \cdot \log^2(1/\epsilon)\right) \\ &\quad = \widetilde{O}\left(w^{1-2\cdot 40^{-d}} \cdot 2^{O(d)} \cdot \log^2(1/\epsilon)\right). \end{aligned}$$

We can assume without loss of generality that $\delta \ge 1/\log n$, as otherwise the seed length in the theorem statement is greater than n. Since $\delta = \frac{1}{4}50^{-d}$, we get

$$50^d \le \frac{1}{4}\log n. \tag{13}$$

Therefore, the $2^{O(d)}$ term in the seed length is at most $\mathrm{polylog}(n)$, so we can absorb it into the $\widetilde{O}$, making both the total seed length and the time complexity of computing each output bit $\widetilde{O}(w^{1-2\cdot 40^{-d}} \cdot \log^2(1/\epsilon))$. Recalling $w = n^{1+\delta}$ and $\epsilon = 2^{-n^\delta}$, the seed length and the time complexity of computing each output bit are bounded by

$$\begin{aligned}
n^{1+3\delta-2\cdot 40^{-d}} \cdot \mathrm{polylog}(n) &\le n^{1-\delta} \cdot n^{-40^{-d}} \cdot \mathrm{polylog}(n) && (40^{-d} > 50^{-d} = 4\delta) \\
&\le n^{1-\delta} \cdot n^{-(50^{-d})^{0.95}} \cdot \mathrm{polylog}(n) && (\frac{\log(40)}{\log(50)} < 0.95) \\
&\le n^{1-\delta} \cdot n^{-\Omega(1/\log^{0.95} n)} \cdot \mathrm{polylog}(n) && \text{(Eq. (13))} \\
&= n^{1-\delta} \cdot 2^{-\Omega(\log^{0.05} n)} \cdot 2^{O(\log\log n)} \\
&\le n^{1-\delta}
\end{aligned}$$

for sufficiently large n. ■

6.4 Implications for MCSP

Let $\mathsf{CC}(f)$ denote the size of the smallest Boolean circuit computing f. In the MCSP problem, we are given an n-bit truth table of a function $f\colon \{0,1\}^{\log n} \to \{0,1\}$ and a size parameter $\theta \in \mathbb{N}$; our job is to determine whether $\mathsf{CC}(f) \le \theta$. In general, the existence of a PRG with a certain "local efficiency" property that fools some model of computation implies that the model cannot compute MCSP [KC00] In particular, we will now show that our PRG for LTF circuits implies that LTF circuits cannot solve MCSP with certain parameters.

In fact, we will show that LTF cirucits cannot even compute a *relaxed* version of MCSP. For two functions $s_1, s_2\colon \mathbb{N} \to \mathbb{N}$ such that $s_1(n) < s_2(n)$, let $\mathsf{gapMCSP}[s_1, s_2]$ denote the promise problem where we are given an n-bit truth table of a function $f\colon \{0,1\}^\ell \to \{0,1\}$, where $n = 2^\ell$, and our goal is to distinguish between the "yes" case $\mathsf{CC}(f) \le s_1(\ell)$ and the "no" case $\mathsf{CC}(f) \ge s_2(\ell)$, under the promise that one of the two holds.

Theorem 6.26 (Lower bound for computing MCSP by LTF circuits)**.** *Let $d\colon \mathbb{N} \to \mathbb{N}$ be a function with $d(n) \le \frac{1}{6}\log\log n$, let $\delta(n) = \frac{1}{8} \cdot 50^{-d(n)}$, and assume that $d(n)$ can be computed in time $O(n^{1-2\delta(n)})$ on a multitape Turing machine. Let $s_1(\ell) = 2^{(1-\delta(2^\ell))\cdot\ell}$, and let $s_2(\ell) = 2^{\ell-1}/\ell$. Then for all sufficiently large ℓ and $n = 2^\ell$, depth-$d(n)$ LTF circuits with $n^{1+\delta(n)}$ wires cannot solve $\mathsf{gapMSCP}[s_1, s_2]$ on truth tables of length n.*

Proof. Fix ℓ and set $n = n(\ell)$, $d = d(n(\ell))$, and $\delta = \delta(n(\ell))$. Let C be a depth-d LTF circuit on n input bits with at most $n^{1+\delta}$ wires. If we pick a function $f\colon \{0,1\}^\ell \to \{0,1\}$ uniformly at random, then with probability at least $\frac{1}{2}$, we have $\mathsf{CC}(f) \ge 2^{\ell-1}/\ell$ [Sha49]. Therefore, if $\mathbb{E}[C(\mathbf{u}_n)] > \frac{1}{2}$, we are done. Assume now that $\mathbb{E}[C(\mathbf{u}_n)] \le \frac{1}{2}$.

Let $G\colon \{0,1\}^s \to \{0,1\}^n$ be the PRG of Theorem 6.25, so $\mathbb{E}[C(G(\mathbf{u}_s))] < 1$, i.e., there is some seed x_* such that $C(G(x_*)) = 0$. The seed length of G and the time to compute each output bit of G on a multitape Turing machine are both bounded by $O(n^{1-2\delta})$ (note that we have defined δ to be smaller by a factor of two than the corresponding parameter in Theorem 6.25). Multitape Turing machines can be simulated by Boolean circuits with logarithmic overhead [PF79], so there is some Boolean circuit A of size $O(n^{1-2\delta}\log n)$ such that for every seed x and every index $i \in [n]$, we have $A(x,i) = G(x)_i$. By hard-wiring the seed x_*, we get a Boolean circuit $B(i) = A(x_*, i) = G(x_*)_i$ of size $O(n^{1-2\delta}\log n)$ whose truth table is $G(x_*)$. Thus, $\mathsf{CC}(G(x_*)) \le O(n^{1-2\delta}\log n)$. Our assumption $d \le \frac{1}{6}\log\log n$ ensures that $50^d \le (\log n)^{1-\Omega(1)}$, and therefore $n^\delta \ge 2^{\log^{\Omega(1)} n} > (\log n)^{\omega(1)}$. Therefore, for sufficiently large ℓ, $\mathsf{CC}(G(x)) \le n^{1-\delta} = 2^{(1-\delta)\ell}$. ■

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. W.M.H. is supported by the NSF GRFP under Grant DGE-1610403 and by a Harrington Fellowship from UT Austin.

References

[Aar16] Scott Aaronson. "$P \stackrel{?}{=} NP$". In: *Open Problems in Mathematics*. Ed. by John Forbes Nash Jr. and Michael Th. Rassias. Springer International Publishing, 2016, pp. 1–122.

[ACW16] Josh Alman, Timothy M. Chan, and Ryan 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.

[And87] Alexander E. Andreev. "On a method for obtaining more than quadratic effective lower bounds for the complexity of π-schemes". In: *Vestnik Moskovskogo Universiteta. Seriya I. Matematika, Mekhanika* 1 (1987), pp. 70–73, 103.

[ASWZ96] Roy Armoni, Michael Saks, Avi Wigderson, and Shiyu Zhou. "Discrepancy sets and pseudorandom generators for combinatorial rectangles". In: *37th Annual Symposium on Foundations of Computer Science (Burlington, VT, 1996)*. IEEE Comput. Soc. Press, Los Alamitos, CA, 1996, pp. 412–421. DOI: 10.1109/SFCS.1996.548500. URL: https://doi.org/10.1109/SFCS.1996.548500.

[AW85] Miklos Ajtai and Avi Wigderson. "Deterministic simulation of probabilistic constant depth circuits". In: *Proc. 26th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 1985.

[BCWZ99] Harry Buhrman, Richard Cleve, Ronald de Wolf, and Christof Zalka. "Bounds for small-error and zero-error quantum algorithms". In: *40th Annual Symposium on Foundations of Computer Science (New York, 1999)*. IEEE Computer Soc., Los Alamitos, CA, 1999, pp. 358–368. DOI: 10.1109/SFFCS.1999.814607. URL: https://doi.org/10.1109/SFFCS.1999.814607.

[BFT98] Harry Buhrman, Lance Fortnow, and Thomas Thierauf. "Nonrelativizing separations". In: *Proc. 13th Annual IEEE Conference on Computational Complexity (CCC)*. 1998, pp. 8–12.

[BNRW07] Harry Buhrman, Ilan Newman, Hein Röhrig, and Ronald de Wolf. "Robust Polynomials and Quantum Algorithms". In: *Theory Comput. Syst.* 40.4 (2007), pp. 379–395. DOI: 10.1007/s00224-006-1313-z. URL: https://doi.org/10.1007/s00224-006-1313-z.

[Bog18] Andrej Bogdanov. "Small Bias Requires Large Formulas". In: *45th International Colloquium on Automata, Languages, and Programming, ICALP 2018, July 9-13, 2018, Prague, Czech Republic*. Ed. by Ioannis Chatzigiannakis, Christos Kaklamanis, Dániel Marx, and Donald Sannella. Vol. 107. LIPIcs. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018, 22:1–22:12. DOI: 10.4230/LIPIcs.ICALP.2018.22. URL: https://doi.org/10.4230/LIPIcs.ICALP.2018.22.

[BR94] Mihir Bellare and John Rompel. “Randomness-efficient Oblivious Sampling”. In: *Proc. 35th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 1994, pp. 276–287.

[BSV14] Eli Ben-Sasson and Emanuele Viola. “Short PCPs with projection queries”. In: *Proc. 41st International Colloquium on Automata, Languages and Programming (ICALP)*. 2014, pp. 163–173.

[CDS19] Eshan Chattopadhyay, Anindya De, and Rocco A. Servedio. “Simple and efficient pseudorandom generators from Gaussian processes”. In: *Proc. 34th Annual IEEE Conference on Computational Complexity (CCC)*. 2019, Art. No. 4, 33.

[CHHL19] Eshan Chattopadhyay, Pooya Hatami, Kaave Hosseini, and Shachar Lovett. “Pseudorandom generators from polarizing random walks”. In: *Theory of Computing* 15 (2019), Paper No. 10, 26.

[CHMY21] Mahdi Cheraghchi, Shuichi Hirahara, Dimitrios Myrisiotis, and Yuichi Yoshida. “One-tape Turing machine and branching program lower bounds for MCSP”. In: *Proc. 38th Symposium on Theoretical Aspects of Computer Science (STACS)*. Vol. 187. LIPIcs. Leibniz Int. Proc. Inform. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2021, Art. 23, 19.

[CJW19] Lijie Chen, Ce Jin, and Richard Ryan Williams. “Hardness Magnification for all Sparse NP Languages”. In: *Proc. 60th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 2019.

[CKKSZ15] Ruiwen Chen, Valentine Kabanets, Antonina Kolokolova, Ronen Shaltiel, and David Zuckerman. “Mining circuit lower bound proofs for meta-algorithms”. In: *Computational Complexity* 24.2 (2015), pp. 333–392.

[CKLM19] Mahdi Cheraghchi, Valentine Kabanets, Zhenjian Lu, and Dimitrios Myrisiotis. “Circuit lower bounds for MCSP from local pseudorandom generators”. In: *Proc. 46th International Colloquium on Automata, Languages and Programming (ICALP)*. Vol. 132. 2019, Art. No. 39, 14.

[CLW20] Lijie Chen, Xin Lyu, and Ryan Williams. “Almost-Everywhere Circuit Lower Bounds from Non-Trivial Derandomization”. In: *Proc. 61st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 2020.

[CMMW19] Lijie Chen, Dylan M. McKay, Cody D. Murray, and R. Ryan Williams. “Relations and Equivalences Between Circuit Lower Bounds and Karp-Lipton Theorems”. In: *Proc. 34th Annual IEEE Conference on Computational Complexity (CCC)*. 2019, 30:1–30:21.

[CSS18] Ruiwen Chen, Rahul Santhanam, and Srikanth Srinivasan. “Average-case lower bounds and satisfiability algorithms for small threshold circuits”. In: *Theory Comput.* 14 (2018), Paper No. 9, 55. DOI: 10.4086/toc.2018.v014a009. URL: https://doi.org/10.4086/toc.2018.v014a009.

[CT19] Lijie Chen and Roei 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.

[CW19] Lijie Chen and R. Ryan 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, 19:1–19:43.

[DGJSV10] Ilias Diakonikolas, Parikshit Gopalan, Ragesh Jaiswal, Rocco A. Servedio, and Emanuele Viola. “Bounded independence fools halfspaces”. In: *SIAM Journal of Computing* 39.8 (2010), pp. 3441–3462.

[DKN10] Ilias Diakonikolas, Daniel M. Kane, and Jelani Nelson. "Bounded independence fools degree-2 threshold functions". In: *Proc. 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 2010, pp. 11–20.

[DORS08] Yevgeniy Dodis, Rafail Ostrovsky, Leonid Reyzin, and Adam Smith. "Fuzzy extractors: how to generate strong keys from biometrics and other noisy data". In: *SIAM J. Comput.* 38.1 (2008), pp. 97–139. ISSN: 0097-5397. DOI: 10.1137/060651380. URL: https://doi.org/10.1137/060651380.

[DRV10] Cynthia Dwork, Guy N. Rothblum, and Salil Vadhan. "Boosting and differential privacy". In: *2010 IEEE 51st Annual Symposium on Foundations of Computer Science—FOCS 2010*. IEEE Computer Soc., Los Alamitos, CA, 2010, pp. 51–60.

[EGLNV98] Guy Even, Oded Goldreich, Michael Luby, Noam Nisan, and Boban Veličković. "Efficient approximation of product distributions". In: *Random Structures Algorithms* 13.1 (1998), pp. 1–16. ISSN: 1042-9832. DOI: 10.1002/(SICI)1098-2418(199808)13:1<1::AID-RSA1>3.0.CO;2-W. URL: https://doi.org/10.1002/(SICI)1098-2418(199808)13:1<1::AID-RSA1>3.0.CO;2-W.

[FK18] Michael A. Forbes and Zander Kelley. "Pseudorandom generators for read-once branching programs, in any order". In: *59th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2018*. IEEE Computer Soc., Los Alamitos, CA, 2018, pp. 946–955. DOI: 10.1109/FOCS.2018.00093. URL: https://doi.org/10.1109/FOCS.2018.00093.

[GHR92] Mikael Goldmann, Johan Håstad, and Alexander Razborov. "Majority gates vs. general weighted threshold gates". In: *Proc. 7th Annual Structure in Complexity Theory Conference*. 1992, pp. 2–13.

[GII+19] Alexander Golovnev, Rahul Ilango, Russell Impagliazzo, Valentine Kabanets, Antonina Kolokolova, and Avishay Tal. "$AC^0[p]$ lower bounds against MCSP via the coin problem". In: *Proc. 46th International Colloquium on Automata, Languages and Programming (ICALP)*. Vol. 132. 2019, Art. No. 66, 15.

[GK98] Mikael Goldmann and Marek Karpinski. "Simulating Threshold Circuits by Majority Circuits". In: *SIAM Journal of Computing* 27.1 (1998), pp. 230–246.

[GKM18] Parikshit Gopalan, Daniel M. Kane, and Raghu Meka. "Pseudorandomness via the discrete Fourier transform". In: *SIAM J. Comput.* 47.6 (2018), pp. 2451–2487. ISSN: 0097-5397. DOI: 10.1137/16M1062132. URL: https://doi.org/10.1137/16M1062132.

[GLSS15] Dmitry Gavinsky, Shachar Lovett, Michael Saks, and Srikanth Srinivasan. "A tail bound for read-k families of functions". In: *Random Structures & Algorithms* 47.1 (2015), pp. 99–108.

[GMRTV12] Parikshit Gopalan, Raghu Meka, Omer Reingold, Luca Trevisan, and Salil Vadhan. "Better pseudorandom generators from milder pseudorandom restrictions". In: *Proc. 53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 2012, pp. 120–129.

[GOWZ10] Parikshit Gopalan, Ryan O'Donnell, Yi Wu, and David Zuckerman. "Fooling functions of halfspaces under product distributions". In: *Proc. 25th Annual IEEE Conference on Computational Complexity (CCC)*. 2010, pp. 223–234.

[GT91] Hans Dietmar Gröger and György" Turán. "On linear decision trees computing Boolean functions". In: *Proc. 18th International Colloquium on Automata, Languages and Programming (ICALP)*. 1991, pp. 707–718.

[GUV09] Venkatesan Guruswami, Christopher Umans, and Salil Vadhan. "Unbalanced expanders and randomness extractors from Parvaresh-Vardy codes". In: *Journal of the ACM* 56.4 (2009), Art. 20, 34.

[GY20] Parikshit Gopalan and Amir Yehudayoff. "Concentration for Limited Independence via Inequalities for the Elementary Symmetric Polynomials". In: *Theory of Computing* 16.17 (2020), pp. 1–29. DOI: 10.4086/toc.2020.v016a017. URL: http://www.theoryofcomputing.org/articles/v016a017.

[Hås14] Johan Håstad. "On the correlation of parity and small-depth circuits". In: *SIAM Journal of Computing* 43.5 (2014), pp. 1699–1708.

[Hås87] Johan Håstad. *Computational Limitations of Small-depth Circuits*. MIT Press, 1987.

[Hås98] Johan Håstad. "The shrinkage exponent of De Morgan formulas is 2". In: *SIAM J. Comput.* 27.1 (1998), pp. 48–64. ISSN: 0097-5397. DOI: 10.1137/S0097539794261556. URL: https://doi.org/10.1137/S0097539794261556.

[HKM12] Prahladh Harsha, Adam Klivans, and Raghu Meka. "An invariance principle for polytopes". In: *Journal of the ACM* 59.6 (2012), 29:1–29:25.

[HLV18] Elad Haramaty, Chin Ho Lee, and Emanuele Viola. "Bounded independence plus noise fools products". In: *SIAM J. Comput.* 47.2 (2018), pp. 493–523. ISSN: 0097-5397. DOI: 10.1137/17M1129088. URL: https://doi.org/10.1137/17M1129088.

[Hoe63] Wassily Hoeffding. "Probability inequalities for sums of bounded random variables". In: *Journal of the American Statistical Association* 58 (1963), pp. 13–30.

[IK17] Russell Impagliazzo and Valentine Kabanets. "Fourier concentration from shrinkage". In: *Computational Complexity* 26.1 (2017), pp. 275–321.

[IMZ19] Russell Impagliazzo, Raghu Meka, and David Zuckerman. "Pseudorandomness from shrinkage". In: *J. ACM* 66.2 (2019), Art. 11, 16. ISSN: 0004-5411. DOI: 10.1145/3230630. URL: https://doi.org/10.1145/3230630.

[IN93] Russell Impagliazzo and Noam Nisan. "The effect of random restrictions on formula size". In: *Random Structures & Algorithms* 4.2 (1993), pp. 121–133.

[INW94] Russell Impagliazzo, Noam Nisan, and Avi Wigderson. "Pseudorandomness for network algorithms". In: *Proc. 26th Annual ACM Symposium on Theory of Computing (STOC)*. 1994, pp. 356–364.

[IPS13] Russell Impagliazzo, Ramamohan Paturi, and Stefan 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.

[IPS93] Russell Impagliazzo, Ramamohan Paturi, and Michael E. Saks. "Size-depth tradeoffs for threshold circuits". In: *Proc. 25th Annual ACM Symposium on Theory of Computing (STOC)*. 1993, pp. 541–550.

[IPZ01] Russell Impagliazzo, Ramamohan Paturi, and Francis Zane. "Which problems have strongly exponential complexity?" In: *Journal of Computer and System Sciences* 63.4 (2001), pp. 512–530.

[Kan11] Daniel 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.

[Kan14] Daniel 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.

[Kan82] R. Kannan. “Circuit-size lower bounds and non-reducibility to sparse sets”. In: *Information and Control* 55.1-3 (1982), pp. 40–56.

[KC00] Valentine Kabanets and Jin-Yi Cai. “Circuit minimization problem”. In: *Proc. 32nd Annual ACM Symposium on Theory of Computing (STOC)*. 2000, pp. 73–79.

[Khr71] V. M. Khrapčenko. “A certain method of obtaining estimates from below of the complexity of π-schemes”. In: *Matematicheskie Zametki* 10 (1971), pp. 83–92.

[KKLMO20] Valentine Kabanets, Sajin Koroth, Zhenjian Lu, Dimitrios Myrisiotis, and Igor 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.

[KL18] Valentine Kabanets and Zhenjian Lu. “Satisfiability and derandomization for small polynomial threshold circuits”. In: *Proc. 22nd International Workshop on Randomization and Approximation Techniques in Computer Science (RANDOM)*. 2018, Art. No. 46, 19.

[KM15] Pravesh K. Kothari and Raghu Meka. “Almost optimal pseudorandom generators for spherical caps”. In: *Proc. 47th Annual ACM Symposium on Theory of Computing (STOC)*. 2015, pp. 247–256.

[KR13] Ilan Komargodski and Ran Raz. “Average-case lower bounds for formula size”. In: *Proc. 45th Annual ACM Symposium on Theory of Computing (STOC)*. 2013, pp. 171–180.

[KRS12] Zohar S. Karnin, Yuval Rabani, and Amir Shpilka. “Explicit dimension reduction and its applications”. In: *SIAM Journal of Computing* 41.1 (2012), pp. 219–249.

[KRT17] Ilan Komargodski, Ran Raz, and Avishay Tal. “Improved average-case lower bounds for De Morgan formula size: matching worst-case lower bound”. In: *SIAM Journal of Computing* 46.1 (2017), pp. 37–57.

[KW16] Daniel M. Kane and Ryan 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.

[Lee19] Chin Ho Lee. “Fourier bounds and pseudorandom generators for product tests”. In: *34th Computational Complexity Conference*. Vol. 137. LIPIcs. Leibniz Int. Proc. Inform. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2019, Art. No. 7, 25.

[LLSZ97] Nathan Linial, Michael Luby, Michael Saks, and David Zuckerman. “Efficient construction of a small hitting set for combinatorial rectangles in high dimension”. In: *Combinatorica* 17.2 (1997), pp. 215–234. ISSN: 0209-9683. DOI: 10.1007/BF01200907. URL: https://doi.org/10.1007/BF01200907.

[LS11] Shachar Lovett and Srikanth 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.

[Lu02] Chi-Jen Lu. “Improved pseudorandom generators for combinatorial rectangles”. In: *Combinatorica* 22.3 (2002), pp. 417–433. ISSN: 0209-9683. DOI: 10.1007/s004930200021. URL: https://doi.org/10.1007/s004930200021.

[MW18] Cody Murray and Ryan 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.

[MZ13] Raghu Meka and David Zuckerman. “Pseudorandom generators for polynomial threshold functions”. In: *SIAM Journal of Computing* 42.3 (2013), pp. 1275–1301.

[Nis92] Noam Nisan. “Pseudorandom generators for space-bounded computation”. In: *Combinatorica* 12.4 (1992), pp. 449–461. ISSN: 0209-9683. DOI: 10.1007/BF01305237. URL: https://doi.org/10.1007/BF01305237.

[Nis93] Noam Nisan. “The communication complexity of threshold gates”. In: *Combinatorics, Paul Erdős is eighty, Vol. 1*. Bolyai Society Mathematical Studies. 1993, pp. 301–315.

[OPS19] Igor Carboni Oliveira, Ján Pich, and Rahul Santhanam. “Hardness magnification near state-of-the-art lower bounds”. In: *Proc. 34th Annual IEEE Conference on Computational Complexity (CCC)*. 2019, Art. No. 27, 29.

[OS18] Igor Carboni Oliveira and Rahul Santhanam. “Hardness Magnification for Natural Problems”. In: *Electronic Colloquium on Computational Complexity: ECCC* 25 (2018), p. 139.

[OST19] Ryan O’Donnell, Rocco A. Servedio, and Li-Yang Tan. “Fooling polytopes”. In: *Proc. 51st Annual ACM Symposium on Theory of Computing (STOC)*. 2019, pp. 614–625.

[PF79] Nicholas Pippenger and Michael J. Fischer. “Relations among complexity measures”. In: *Journal of the ACM* 26.2 (1979), pp. 361–381.

[PZ93] Michael S. Paterson and Uri Zwick. “Shrinkage of De Morgan formulae under restriction”. In: *Random Structures & Algorithms* 4.2 (1993), pp. 135–150.

[Rei11] Ben W. Reichardt. “Reflections for quantum query algorithms”. In: *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, Philadelphia, PA, 2011, pp. 560–569.

[ROS94] V. P. Roychowdhury, A. Orlitsky, and Kai-Yeung Siu. “Lower bounds on threshold and related circuits via communication complexity”. In: *IEEE Transactions on Information Theory* 40.2 (1994), pp. 467–474.

[RR97] Alexander A. Razborov and Steven Rudich. “Natural proofs”. In: *Journal of Computer and System Sciences* 55.1, part 1 (1997), pp. 24–35.

[RS10] Yuval Rabani and Amir Shpilka. “Explicit Construction of a Small epsilon-Net for Linear Threshold Functions”. In: *SIAM Journal of Computing* 39.8 (2010), pp. 3501–3520.

[San09] Rahul Santhanam. “Circuit lower bounds for Merlin-Arthur classes”. In: *SIAM Journal of Computing* 39.3 (2009), pp. 1038–1061.

[San10] Rahul 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.

[Ser07] Rocco A. Servedio. “Every linear threshold function has a low-weight approximator”. In: *Computational Complexity* 16.2 (2007), pp. 180–209.

[Sha49] Claude E. Shannon. “The synthesis of two-terminal switching circuits”. In: *Bell System Tech. J.* 28 (1949), pp. 59–98. ISSN: 0005-8580. DOI: 10.1002/j.1538-7305.1949.tb03624.x. URL: https://doi.org/10.1002/j.1538-7305.1949.tb03624.x.

[She13] Alexander A. Sherstov. “Making polynomials robust to noise”. In: *Theory Comput.* 9 (2013), pp. 593–615. DOI: 10.4086/toc.2013.v009a018. URL: https://doi.org/10.4086/toc.2013.v009a018.

[Sho90] Victor Shoup. “New algorithms for finding irreducible polynomials over finite fields”. In: *Mathematics of Computation* 54.189 (1990), pp. 435–447.

[SSTT16] Takayuki Sakai, Kazuhisa Seto, Suguru Tamaki, and Junichi 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.

[ST17a] Rocco Servedio and Li-Yang Tan. “Learning and fooling depth-two threshold circuits”. Unpublished manuscript. 2017.

[ST17b] Rocco A. Servedio and Li-Yang Tan. “Fooling intersections of low-weight halfspaces”. In: *Proc. 58th Annual IEEE Conference on Computational Complexity (CCC)*. 2017, pp. 824–835.

[ST18] Rocco A. Servedio and Li-Yang Tan. “Luby-Veličković-Wigderson revisited: improved correlation bounds and pseudorandom generators for depth-two circuits”. In: *Approximation, randomization, and combinatorial optimization. Algorithms and techniques*. Vol. 116. LIPIcs. Leibniz Int. Proc. Inform. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2018, Art. No. 56, 20.

[ST19] Rocco A. Servedio and Li-Yang Tan. “Improved pseudorandom generators from pseudorandom multi-switching lemmas”. In: *Proc. 23rd International Workshop on Randomization and Approximation Techniques in Computer Science (RANDOM)*. 2019, Art. No. 45, 23.

[Sub61] B. A. Subbotovskaja. “Realization of linear functions by formulas using $\vee$, &, $-$”. In: *Soviet Mathematics. Doklady* 2 (1961), pp. 110–112.

[SW13] Rahul Santhanam and Ryan Williams. “On medium-uniformity and circuit lower bounds”. In: *Proc. 28th Annual IEEE Conference on Computational Complexity (CCC)*. 2013, pp. 15–23.

[Tal14] Avishay 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. DOI: 10.1109/FOCS.2014.65. URL: https://doi.org/10.1109/FOCS.2014.65.

[Tal17a] Avishay Tal. “Formula lower bounds via the quantum method”. In: *Proc. 49th Annual ACM Symposium on Theory of Computing (STOC)*. 2017, pp. 1256–1268.

[Tal17b] Avishay Tal. “Tight Bounds on the Fourier Spectrum of AC0”. In: *Proc. 32nd Annual IEEE Conference on Computational Complexity (CCC)*. 2017, 15:1–15:31.

[Tam16] Suguru Tamaki. “A Satisfiability Algorithm for Depth Two Circuits with a Sub-Quadratic Number of Symmetric and Threshold Gates”. In: *Electronic Colloquium on Computational Complexity: ECCC* 23 (2016), p. 100.

[Tel18] Roei Tell. “Quantified Derandomization of Linear Threshold Circuits”. In: *Proc. 50th Annual ACM Symposium on Theory of Computing (STOC)*. 2018, pp. 855–865.

[Vad12] Salil P. Vadhan. *Pseudorandomness*. Foundations and Trends in Theoretical Computer Science. Now Publishers, 2012.

[Vio07] Emanuele Viola. “Pseudorandom bits for constant-depth circuits with few arbitrary symmetric gates”. In: *SIAM Journal of Computing* 36.5 (2007), pp. 1387–1403.

[Vio09] Emanuele Viola. “The sum of d small-bias generators fools polynomials of degree d”. In: *Computational Complexity* 18.2 (2009), pp. 209–217.

[Vio14] Emanuele Viola. “Randomness Buys Depth for Approximate Counting”. In: *Computational Complexity* 23.3 (2014), pp. 479–508.

[Vio15] Emanuele Viola. "The communication complexity of addition". In: *Combinatorica* 35.6 (2015), pp. 703–747. ISSN: 0209-9683. DOI: 10.1007/s00493-014-3078-3. URL: https://doi.org/10.1007/s00493-014-3078-3.

[Wil11] Ryan Williams. "Non-uniform ACC circuit lower bounds". In: *Proc. 26th Annual IEEE Conference on Computational Complexity (CCC)*. 2011, pp. 115–125.

[Wil13] Ryan Williams. "Improving Exhaustive Search Implies Superpolynomial Lower Bounds". In: *SIAM Journal of Computing* 42.3 (2013), pp. 1218–1244.

[Wil18] Richard Ryan Williams. "New algorithms and lower bounds for circuits with linear threshold gates". In: *Theory of Computing* 14 (2018), Paper No. 17, 25.

A An improved low-error PRG for formulas of LTFs

A.1 Low-error approximating polynomials for De Morgan formulas

Recall that for a Boolean function $f\colon \{0,1\}^n \to \{0,1\}$, the *$\delta$-error approximate degree* $\widetilde{\deg}_\delta(f)$ is the smallest degree of a polynomial p such that for all x, $|f(x) - p(x)| \leq \delta$. We rely on a beautiful result from quantum computing by Reichardt [Rei11].

Theorem A.1 ([Rei11])**.** *If F is a size-s De Morgan formula, then $\widetilde{\deg}_{1/3}(F) \leq O(\sqrt{s})$.*

While Reichardt's result gives a polynomial of degree $\approx \sqrt{s}$ that pointwise approximates f to error $1/3$, it is desirable in many cases to reduce the error much further to some small parameter δ (that may depend on s). It is well-known (see [BNRW07]) that error reduction can be done in a black-box fashion, yielding a polynomial of degree $O(\sqrt{s}\log(1/\delta))$ that point-wise approximates the formula up to error δ. However, this result is not tight, and in the next theorem we show that one can get a better dependency on the error parameter, namely $\sqrt{s\log(1/\delta)}$. Such a result was known for the special case of the OR function on s variables, and is tight for this special case for any $\delta > 2^{-s}$ [BCWZ99].

Theorem A.2. *If F is a size-s De Morgan formula, then $\widetilde{\deg}_\delta(F) \leq O(\sqrt{s\log(1/\delta)})$.*

Proof. Take a formula for F of size s. Let $k = \log(1/\delta)$. By Lemma 5.3, we can write F as a composition of a read-once top formula T of size $m = O(k)$ and bottom-formulas $B_1, \ldots, B_m$ each of size at most s/k.

By Theorem A.1, we know that each B_i can be point-wise approximated by a polynomial q_i of degree at most $O(\sqrt{s/k})$.

Take the unique multilinear polynomial p that computes the formula T exactly. $\deg(T) \leq m$ since T is a formula with m leaves. Take the robust version p_{robust} of p with parameter δ, as guaranteed from Theorem 6.14. By construction $\deg(p_{\text{robust}}) = O(k + \log(1/\delta)) = O(k)$. Consider the composition $f(x) = p_{\text{robust}}(q_1(x), q_2(x) \ldots, q_m(x))$. The polynomial f is of degree at most

$$O(k) \cdot O(\sqrt{s/k}) = O(\sqrt{sk}) = O\left(\sqrt{s\log(1/\delta)}\right).$$

Moreover, on any input $x \in \{0,1\}^n$ we have

$$(q_1(x), q_2(x) \ldots, q_m(x)) = (B_1(x), \ldots, B_m(x)) + \eta$$

where $\eta \in [-1/3, 1/3]^m$, and thus Theorem 6.14 gives us

$$|f(x) - F(x)| = |p_{\text{robust}}(q_1(x), \ldots, q_m(x)) - p(B_1(x), \ldots, B_m(x))| \leq \delta. \quad \blacksquare$$

A.2 PRGs from approximating polynomials

Bounds on approximate degree can be combined with Theorem 6.13:

Theorem A.3. *Let* $\mathbf{x}$ *be a distribution over* $(\{0,1\}^{n/k})^k$ *that* ϵ*-fools* k*-dimensional combinatorial rectangles. Let* $f(x) = F(g_1(x),\dots,g_s(x))$*, where* $\widetilde{\deg}_\delta(F) \le d$ *and* $g_1,\dots,g_s\colon (\{0,1\}^{n/k})^k \to \{0,1\}$ *can be computed by randomized* k*-party number-in-hand communication protocols with communication cost* R *and failure probability* $1/6$*. Then* $\mathbf{x}$ *fools* f *with error*

$$2\delta + (\epsilon \cdot (4s+4)^d)^{\Omega(1/R)} \cdot 2^{O(d)}.$$

Proof. Let p be a degree-d polynomial such that for all $x \in \{0,1\}^s$, $|p(x) - F(x)| \le \delta$. Define $q(x) = \frac{p(x)}{1+\delta}$, so that q takes values in $[0,1]$ and we still have $|q(x) - F(x)| \le 2\delta$. By Theorem 6.13, $\mathbf{x}$ fools $q(g_1(x),\dots,g_s(x))$ with error $(\epsilon \cdot L_1(q))^{\Omega(1/R)} \cdot 2^{O(d)}$. To bound $L_1(q)$, it is helpful to move to $\{\pm 1\}$ inputs, i.e., define

$$r(x_1,\dots,x_s) = q\left(\frac{1}{2} - \frac{1}{2}x_1,\dots,\frac{1}{2} - \frac{1}{2}x_s\right),$$

so that

$$q(x_1,\dots,x_s) = r(1-2x_1,\dots,1-2x_s).$$

By considering each monomial individually, we see that $\deg(r) \le \deg(q) = d$ and $L_1(q) \le L_1(r) \cdot 4^d$. We may think of r as the Fourier expansion of a function $r\colon \{\pm 1\}^s \to [0,1]$. This shows that each individual coefficient of r has absolute value at most 1. The number of coefficients is at most $(s+1)^d$, so $L_1(r) \le (s+1)^d$. ∎

Now we are ready to give our improved low-error PRG for size-s De Morgan formulas with LTFs at the leaves. Let $\mathtt{FORMULA}[s]$ denote the class of De Morgan formulas of size s.

Corollary A.4. *For all* $n, s \in \mathbb{N}$ *and all* $\epsilon > 0$*, there is a* $\mathrm{poly}(n)$*-time computable PRG for* $\mathtt{FORMULA}[s] \circ \mathtt{LTF}$ *with output length* n *and seed length*

$$\widetilde{O}(n^{1/2}s^{1/4}\log^{1/4}(1/\epsilon) + n^{1/2}\log^{1/2}(1/\epsilon)).$$

Proof. By Theorem A.2, Theorem A.3, and Theorem 6.15, for any k, it suffices to fool k-dimensional combinatorial rectangles with error

$$2^{-\widetilde{O}(\sqrt{s\log(1/\epsilon)}\cdot k\log n)}.$$

By Theorem 6.18, this can be done with seed length

$$\widetilde{O}\left(n/k + \sqrt{s\log(1/\epsilon)}\cdot k\log n\right).$$

To balance the two terms, we choose $k = n^{1/2}s^{-1/4}\log^{-1/4}(1/\epsilon)$. ∎

B An improved low-error PRG for branching programs and general formulas

In this appendix we construct PRGs with good dependency on the error of size-s branching programs and for size-s formulas over an arbitrary basis. Impagliazzo, Meka, and Zuckerman showed how to fool these classes with seed length of $s^{1/2+o(1)}$ and error $1/\mathrm{poly}(s)$ (for $s \ge n$) [IMZ19]. We will achieve an improved seed length of $\sqrt{s}\cdot\mathrm{polylog}(n/\epsilon)$.

Theorem B.1 (Low-error PRG for branching programs and formulas over a general basis)**.** *For any $n, s \in \mathbb{N}$ and any $\epsilon > 0$, there is an ϵ-PRG for size-s branching programs and size-s formulas over an arbitrary basis with output length n and seed length $\sqrt{s} \cdot \mathrm{polylog}(n/\epsilon)$, computable in time $\mathrm{poly}(n)$.*

The main improvement is the better dependence on ϵ; our PRG remains nontrivial even when the error parameter is $2^{-s^{\Omega(1)}}$. Even for constant error, our seed length is superior in terms of the lower order terms. The $s^{o(1)}$ term in the IMZ seed length is, more specifically, $2^{O(\sqrt{\log n})}$. Our seed length replaces $2^{O(\sqrt{\log n})}$ with $\mathrm{polylog}(n)$.

B.1 Shrinkage of all shifts simultaneously

We will present the proof for branching programs; the proof for formulas over a general basis is essentially identical. Like the PRG for De Morgan formulas, the PRG is based on the phenomenon that branching programs *shrink* under random restrictions. In Impagliazzo, Meka, and Zuckerman's work [IMZ19], the bottleneck preventing them from achieving low error is that the chance of a branching program failing to shrink is too high.

To evade this obstacle, one approach would be to use a hybrid decision tree model like we did with De Morgan formulas. We will take a different approach that gives better parameters and is simpler in some respects. The key observation is that when shrinkage does occur, *all shifts* of the branching program shift as well. (This is not necessarily true of De Morgan formulas.)

To state this precisely, for a function $f\colon \{0,1\}^n \to \{0,1\}$ and a string $y \in \{0,1\}^n$, define $f^{\oplus y}(x) = f(x \oplus y)$. Let $\mathrm{BP}(f)$ be the size of the smallest branching program computing f.

Lemma B.2 (Shrinkage of all shifts of branching programs)**.** *Let $p \in (0,1)$ and let $\boldsymbol{\rho}$ be p-regular restriction. Then for any size-s branching program f,*

$$\Pr[\forall y, \mathrm{BP}((f^{\oplus y})\restriction_{\boldsymbol{\rho}}) \leq 4ps] \geq 3/4.$$

Proof. For $i \in [n]$, let s_i be the number of nodes in f that query x_i. Define a random variable $\mathbf{a}_i$ by

$$\mathbf{a}_i = \begin{cases} s_i & \text{if } \boldsymbol{\rho}_i = \star \\ 0 & \text{if } \boldsymbol{\rho}_i \neq \star. \end{cases}$$

Then $\mathbb{E}[\sum_i \mathbf{a}_i] \leq ps$, so by Markov's inequality, $\Pr[\sum_i \mathbf{a}_i \leq 4ps] \geq 3/4$. Now, fix any shift $y \in \{0,1\}^n$. The function $(f^{\oplus y})\restriction_{\boldsymbol{\rho}}$ can be computed by a branching program of size $\sum_i \mathbf{a}_i$ obtained from f by deleting all vertices v that query some x_i with $\rho_i \neq \star$. The incoming edges to such a vertex are redirected to the vertex reached from v by following the edge labeled $\boldsymbol{\rho}_i \oplus y_i$. ∎

B.2 A good set of restrictions with high probability

The first step of our PRG is to independently sample pseudorandom restrictions $\boldsymbol{\rho}_1, \ldots, \boldsymbol{\rho}_t \in \{0,1,\star\}^n$. Ideally, we would like these restrictions to shrink all shifts of f like in Lemma B.2, and we additionally would like the $\star$ coordinates of these restrictions to collectively cover almost all of $[n]$. That won't necessarily happen, but we now will argue that with high probability, some *subset* of the restrictions satisfies both conditions. This is how we are able to tolerate the high failure probability in Lemma B.2.

Lemma B.3. *Let f be a size-s branching program on n input bits, and let $p, \epsilon > 0$. For $t = O(p^{-1} \log n + \log(1/\epsilon)$, sample independent restrictions $\boldsymbol{\rho}_1, \ldots, \boldsymbol{\rho}_t \in \{0,\star\}^n$, where each $\boldsymbol{\rho}_i$ is p-regular and 4-wise independent. Then except with probability ϵ, there is some set $\mathbf{J} \subseteq [t]$ such that*

1. *For every* $j \in \mathbf{J}$*, for every* $y \in \{0,1\}^n$, $\mathrm{BP}((f^{\oplus y})|_{\boldsymbol{\rho}^{(j)}}) \leq 4ps$.

2. $\left|\bigcup_{j\in\mathbf{J}} \boldsymbol{\rho}_j^{-1}(\star)\right| \geq n - O(p^{-1})$.

Proof. The proof is very similar to the proof of Claim 5.9. Suppose we have already sampled $\boldsymbol{\rho}_1, \ldots, \boldsymbol{\rho}_{j-1}$, and we have already defined $\mathbf{J} \cap [j-1]$. Let $\mathbf{I}$ be the set of coordinates in $[n]$ that are not yet covered by a $\star$, i.e.,

$$\mathbf{I} = [n] \setminus \left(\bigcup_{i\in\mathbf{J}\cap[j-1]} \boldsymbol{\rho}_i^{-1}(\star) \right).$$

Let us include $j \in \mathbf{J}$ if $|\boldsymbol{\rho}_j^{-1}(\star) \cap \mathbf{I}| \geq p|\mathbf{I}|/2$ and for all y, $\mathrm{BP}((f^{\oplus y})|_{\boldsymbol{\rho}_j}) \leq 4ps$. We shall also include $j \in \mathbf{J}$ if $|\mathbf{I}| < 2400p^{-1}$. If $|\mathbf{I}| \geq 2400p^{-1}$, then by Theorem 4.5 with $\Delta = p|\mathbf{I}|/2$,

$$\Pr_{\boldsymbol{\rho}_j}[|\boldsymbol{\rho}_j^{-1}(\star) \cap \mathbf{I}| \leq p|\mathbf{I}|/2] \leq \frac{1}{16}.$$

Combining with Lemma B.2, we see that conditioned on any values of $\boldsymbol{\rho}_1, \ldots, \boldsymbol{\rho}_j$, we have

$$\Pr_{\boldsymbol{\rho}_j}[j \in \mathbf{J}] \geq 1 - \frac{1}{4} - \frac{1}{16} > \frac{2}{3}.$$

This implies that after picking $\boldsymbol{\rho}_1, \ldots, \boldsymbol{\rho}_t$, with high probability, $|\mathbf{J}| \geq \Omega(t)$. Indeed, by Azuma's inequality [DRV10, Lemma B.1], $|\mathbf{J}| \geq t/2$ except with probability $e^{-\Omega(t)} \leq \epsilon$.

Suppose now that $|\mathbf{J}| \geq t/2$. We start with n coordinates not covered by $\star$, and whenever we include $j \in \mathbf{J}$, either we are down to $2400p^{-1}$ coordinates not covered, or else the number of coordinates not covered decreases by a factor of $(1 - p/2)$. If the latter case happens every time, then every coordinate is covered, because $(1 - p/2)^{t/2} \cdot n < 1$ (here we use $t \geq \Omega(p^{-1} \log n)$). Therefore, either way we must eventually cover all but $2400p^{-1}$ coordinates. ∎

B.3 The improved analysis of the IMZ generator

Now we shall describe the full PRG construction. The construction is similar to our PRG for De Morgan formulas (Section 5) (and the original PRG by Impagliazzo, Meka, and Zuckerman [IMZ19]).

Sample $\boldsymbol{\rho}_1, \ldots, \boldsymbol{\rho}_t$ as in Lemma B.3, with $p = 1/\sqrt{s}$.[29] Let $G\colon \{0,1\}^d \to \{0,1\}^n$ be a $(4\sqrt{s})$-wise independent generator. Let $\mathsf{Ext}\colon \{0,1\}^m \times \{0,1\}^{d_{\mathsf{Ext}}} \to \{0,1\}^d$ be a $(k_{\mathsf{Ext}}, \epsilon_{\mathsf{Ext}})$-extractor (the values of these parameters will be specified later.) Sample independent uniform random strings $\mathbf{a} \in \{0,1\}^m$ and $\mathbf{y}_1, \ldots, \mathbf{y}_t \in \{0,1\}^{d_{\mathsf{Ext}}}$. Let $\mathbf{z}_i = \mathsf{Ext}(\mathbf{a}, \mathbf{y}_i)$. Furthermore, sample an $O(\sqrt{s})$-wise independent string $\mathbf{w} \in \{0,1\}^n$. Our PRG outputs

$$\mathbf{w} \oplus \bigoplus_{i=1}^{t} G(\mathbf{z}_i) \circ \rho_i.$$

The proof of correctness is once again a hybrid argument. However, a key difference is that in the hybrid argument, we will consider *fixed* restrictions. In particular, let f be a size-s branching program, and consider some fixed $\rho_1, \ldots, \rho_t$ and some set $J \subseteq [t]$ satisfying the conclusion of Lemma B.3. Let $t' = |J|$, and let σ be a permutation on $[t]$ such that $J = \{\sigma(1), \ldots, \sigma(t')\}$.

[29] Actually, to optimize the low order terms in the seed length, we should use a value of p that is larger by a factor of $\sqrt{\log(n/\epsilon)}$. We use $p = 1/\sqrt{s}$ just for simplicity.

Define

$$\mathbf{e} = \mathbf{w} \oplus \bigoplus_{i \notin J} G(\mathbf{z}_i) \circ \rho_i.$$

Sample independent uniform strings $\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(t')}$, and define hybrid distributions $\mathbf{h}_0, \ldots, \mathbf{h}_{t'}$ by

$$\mathbf{h}_j = \mathbf{e} \oplus \left(\bigoplus_{i=1}^{j} (\mathbf{u}^{(j)} \circ \rho_{\sigma(i)}) \right) \oplus \left(\bigoplus_{i=j+1}^{t'} (G(\mathbf{z}_{\sigma(i)}) \circ \rho_{\sigma(i)}) \right).$$

We stress that $\rho_1, \ldots, \rho_t$ and J are fixed at this point, so the randomness of $\mathbf{h}_j$ is due to the random choices of $\mathbf{w}, \mathbf{a}, \mathbf{y}_1, \ldots, \mathbf{y}_t$.

Claim B.4. *For each $j \in [t']$,*

$$\mathbb{E}[f(\mathbf{h}_{j-1})] - \mathbb{E}[f(\mathbf{h}_j)] \leq 3\epsilon_{\mathsf{Ext}}.$$

Proof. Define $\mathbf{e}'$ to be the random variable such that

$$\mathbf{h}_{j-1} = \mathbf{e}' \oplus (G(\mathbf{z}_{\sigma(j)}) \circ \rho_{\sigma(j)})$$
$$\mathbf{h}_j = \mathbf{e}' \oplus (\mathbf{u}^{(j)} \circ \rho_{\sigma(j)}).$$

Note that $\mathbf{e}'$ is *not* independent of $\mathbf{a}$ or $\mathbf{w}$, but $\mathbf{e}'$ is nevertheless independent of $\mathbf{y}_j$. Let $\mathbf{F} = (f^{\oplus \mathbf{e}'})|_{\rho_{\sigma(j)}}$, so $\mathbf{F}$ (as a function) is a random variable (since it depends on $\mathbf{e}'$). Let us bound the support size of that random variable.

Since $\sigma(j) \in J$, we have $\Pr_{\mathbf{e}'}[\mathrm{BP}(\mathbf{F}) \leq 4\sqrt{s}] = 1$. Therefore, $\mathbf{F}$ can be described using $O(\sqrt{s} \log n)$ bits. As in the proof of Claim 5.8, this implies that $\widetilde{H}_\infty(\mathbf{a} \mid \mathbf{F}) \geq m - O(\sqrt{s} \log n)$. Setting k_{Ext} to be the same value $m = O(\sqrt{s} \log n)$, this implies that $(\mathbf{F}, \mathbf{z}_{\sigma(j)}) \sim_{3\epsilon_{\mathsf{Ext}}} (\mathbf{F}, \mathbf{u})$, where $\mathbf{u}$ is a uniform random d-bit string independent of $\mathbf{F}$. It follows that

$$|\mathbb{E}[\mathbf{F}(G(\mathbf{z}_{\sigma(j)}))] - \mathbb{E}[\mathbf{F}(G(\mathbf{u}))]| \leq 3\epsilon_{\mathsf{Ext}}.$$

Since $\mathrm{BP}(\mathbf{F}) \leq 4\sqrt{s}$, G perfectly fools $\mathbf{F}$, so

$$|\mathbb{E}[\mathbf{F}(G(\mathbf{z}_{\sigma(j)})) - \mathbb{E}[\mathbf{F}(\mathbf{u}^{(j)})] \leq 3\epsilon_{\mathsf{Ext}}.$$

Since $\mathbf{F}(G(\mathbf{z}_{\sigma(j)})) = f(\mathbf{h}_{j-1})$ and $\mathbf{F}(\mathbf{u}^{(j)}) = f(\mathbf{h}_j)$, this completes the proof. ∎

Proof of Theorem B.1. By the triangle inequality,

$$|\mathbb{E}[f(\mathbf{h}_0)] - \mathbb{E}[f(\mathbf{h}_{t'})]| \leq 3t'\epsilon_{\mathsf{Ext}} \leq 3t\epsilon_{\mathsf{Ext}}.$$

Now, $\mathbf{h}_0$ is our pseudorandom distribution. Meanwhile, we claim that $\mathbf{h}_{t'}$ is a perfectly uniform random string. Indeed, observe that

$$(\mathbf{h}_{t'})_i = \begin{cases} \mathbf{e}_i & \text{if } i \notin \bigcup_{j \in J} \rho^{-1}(\star) \\ \text{a fresh uniform random bit} & \text{otherwise.} \end{cases}$$

Since we are assuming the conclusion of Lemma B.3, the number of coordinates i with $i \notin \bigcup_{j \in J} \rho^{-1}(\star)$ is at most $O(\sqrt{s})$. The random variable $\mathbf{w}$ ensures that $\mathbf{e}$ is uniform random on those coordinates. Thus, taking into account the failure probability in Lemma B.3, the total error of our pseudorandom generator is at most $\epsilon + 3t\epsilon_{\mathsf{Ext}} = O(\epsilon)$, choosing $\epsilon_{\mathsf{Ext}} = \epsilon/t$.

Finally, let us bound the seed length of our PRG. Each $\boldsymbol{\rho}_i$ costs $O(\log n)$ truly random bits. The generator G has seed length $d = O(\sqrt{s} \log n)$. Take Ext to be the GUV extractor [GUV09]. Then it suffices to take $m = O(\sqrt{s} \log n)$ and $d_{\mathsf{Ext}} = O(\log(n/\epsilon))$. Summing up gives the claimed seed length. ∎